```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      June 13, 2001, 20:21:23 ; Search time 46.05 Seconds
                              (without alignments)
                              108.942 Million cell updates/sec

Title:       US-09-331-930A-2
Perfect score: 394
Sequence:    1 MIEVVCNDRLGKKVRVKCNT.........HVSLGDYEIHDGMNLELYYQ 73

Scoring table: BLOSUM62
             Gapop 10.0 , Gapext 0.5

rched:      198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters:   198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    PIR_67:*
          1:  pir1:*
          2:  pir2:*
          3:  pir3:*
          4:  pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	331	84.0	73	2	T25763	hypothetical prote
2	306	77.7	73	2	T47458	hypothetical prote
3	301	76.4	73	2	T40200	ubiquitin-like pro
4	252	64.0	73	2	S78735	protein YNR032c-a
5	80	20.3	631	2	S55243	upiquitin-like pro
6	70.5	17.9	299	2	S60529	envelope polyprote
7	70.5	17.9	299	2	S60528	envelope polyprote
8	69	17.5	226	2	T29404	hypothetical prote
9	68.5	17.4	831	2	A70363	mannose-1-phosphat
10	68	17.3	294	2	S60545	envelope polyprote
11	68	17.3	294	2	S60524	envelope polyprote
12	66	16.8	303	2	S60550	envelope polyprote
13	66	16.8	303	2	S60549	envelope polyprote
14	65.5	16.6	504	2	T52254	cytochrome P450 [i
15	65.5	16.6	797	2	D70581	probable cation-tr
16	65.5	16.5	333	2	B64380	hypothetical prote
17	64.5	16.4	165	2	T12774	hypothetical yokJ
18	64	16.2	838	2	T47828	hypothetical prote
19	63.5	16.1	374	2	T02315	envelope polyprote
20	62	15.7	297	2	S60538	envelope polyprote
21	62	15.7	362	2	T34432	hypothetical prote
22	62	15.7	527	2	S42512	recombination-acti
23	62	15.7	662	2	S55274	alpha-L-arabinofur
24	61.5	15.6	276	2	H70869	probable Enoyl-CoA
25	61	15.5	71	1	S42911	env polyprotein -
26	61	15.5	177	2	W2VZA9	20K HindIII-C prot
27	61	15.5	299	2	S60523	envelope polyprote
28	61	15.5	575	2	F69432	fumarate reductase
29	60.5	15.4	339	1	LUCH2	annexin II - chick
30		15.4	346	2	S55403	cymA protein precu
31		15.4	1275	2	A38985	nucleotide exchang
32		15.2	69	2	A25161	tray protein - Sal
33		15.2	71	2	S42908	env polyprotein -
34		15.2	71	2	S42912	env polyprotein -
35		15.2	71	2	S42898	env polyprotein -
36		15.2	71	2	S42901	env polyprotein -
37		15.2	84	2	S76443	hypothetical prote
38		15.2	214	2	H82745	acyl-[ACP]-UDP-N-a
39		15.2	563	2	F81888	probable D-lactate
40		15.2	563	2	D81134	D-lactate dehydrog
41		15.2	741	1	S20923	photosystem I prot
42		15.1	77	2	S35846	envelope protein -
43		15.1	242	2	T32806	hypothetical prote
44		15.1	261	2	E69455	conserved hypothet
45		15.1	497	2	C75618	hypothetical prote

ALIGNMENTS

```
RESULT    1
T25763
hypothetical protein F46F11.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T25763
R;Pauley, A.; Gattung, S.
submitted to the EMBL Data Library, February 1997
A;Description: The sequence of C. elegans cosmid F46F11.
A;Reference number: Z20083
A;Accession: T25763
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-73 <PAU>
A;Cross-references: EMBL:U88173; PIDN:AAB42266.1; GSPDB:GN00019; CESP:F46F11.4
A;Experimental source: strain Bristol N2; clone F46F11
C;Genetics:
C;Gene: CESP:F46F11.4
A;Map position: 1
A;Introns: 38/2

     Query Match             84.0%;  Score 331;  DB 2;  Length 73;
     Best Local Similarity   80.8%;  Pred. No. 1.8e-30;
     Matches   59;  Conservative   6;  Mismatches   8;  Indels   0;  Gaps   0;

Qy      1 MIEVVCNDRLGKKVRVKCNTDDFIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDY 60
          |||: |||||||||:||| |||||||| |||||||||| |||||||||||::|||||:|
Db      1 MIEITVNDRLGKKVRIKCNPSDFIGDLKKLIAAQTGTRWEKIVLKKWYTIYKDHITLMDY 60

Qy     61 EIHDGMNLELYYQ 73
          |||:| | |||||
Db     61 EIHEGFNFELYYQ 73


RESULT    2
T47458
hypothetical protein T14D3.120 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47458
R;Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Lemcke,
submitted to the Protein Sequence Database, February 2000
A;Reference number: 224467
A;Accession: T47458
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-73 <JOR>
A;Cross-references: EMBL:AL138649
A;Experimental source: cultivar Columbia; BAC clone T14D3
C;Genetics:
A;Map position: 3
```

A;Note: T14D3.120

Query Match   77.7%;   Score 306;   DB 2;   Length 73;
Best Local Similarity   79.2%;   Pred. No. 1.2e-27;
Matches   57;   Conservative   6;   Mismatches   9;   Indels   0;   Gaps   0;

```
Qy    1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDY 60
          |||||  ||||||||||||||| :||||||||||| || ::|||: ||||||
Db    1 MIEVVLNDRLGKKVRVKCNEEDTIGDLKKLVAAQTGTRPEKIRIQKWYNIYKDHIPLKDY 60

Qy   61 EIHDGMNLELYY 72
          |||||| |||||
Db   61 EIHDGMGLELYY 72
```

RESULT   3
T40200
ubiquitin-like protein - fission yeast (Schizosaccharonyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
A;Accession: T40200
R;Oliver, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, March 1999
A;Reference number: Z21912
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-73 <OLI>
A;Cross-references: EMBL:AL049190; PIDN:CAB39137.1; GSPDB:GN00067; SPDB:SPBC31E1.03
C;Genetics:
A;Gene: SPDB:SPBC31E1.03
A;Map position: 2
A;Introns: 6/2; 40/2

Query Match   76.4%;   Score 301;   DB 2;   Length 73;
Best Local Similarity   73.6%;   Pred. No. 4.3e-27;
Matches   53;   Conservative   12;   Mismatches   7;   Indels   0;   Gaps   0;

```
Qy    1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDY 60
          ||||:||||||||||||| ||||||||||| :|||||||  :|||::|||:: || ||
Db    1 MIEVLCNDRLGKKVRVKCMPDDTVGDFKKLVAAQTGTGTDPRRIVLKKWHSVFKDNITLADY 60

Qy   61 EIHDGMNLELYY 72
          |||||:|:||
Db   61 EIHDGMSLEMY 72
```

RESULT   4
735
protein YNR032c-a - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 15-Jan-1999 #sequence_revision 15-Jan-1999 #text_change 15-Jan-1999
C;Accession: S78735
R;Pohl, T.M.
submitted to the Protein Sequence Database, April 1996
A;Reference number: S63346
A;Accession: S78735
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-73 <POH>
A;Cross-references: EMBL:Z71647; MIPS:YNR032c-a
A;Experimental source: strain S288C
C;Genetics:
A;Map position: 14R

Query Match   64.0%;   Score 252;   DB 2;   Length 73;
Best Local Similarity   65.3%;   Pred. No. 1.4e-21;
Matches   47;   Conservative   12;   Mismatches   13;   Indels   0;   Gaps   0;

```
Qy    1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDY 60
          ||||  |||||||||||| |||||||||||| | |||||||| :: ||||||||::|| ||
Db    1 MIEVVNDRLGKKVRVKCLAEDSVGDFKKVLSLQIGTQPNKIVLQKGGSVLKDHISLEDY 60

Qy   61 EIHDGMNLELYY 72
          :|| ||||||
Db   61 EVHDQTNLELYY 72
```

RESULT   5
S55243
upiquitin-like protein 8 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 24-Sep-1999
C;Accession: S55243; S61068
R;Callis, J.; Carpenter, T.; Sun, C.W.; Vierstra, R.D.
Genetics 139, 921-939, 1995
A;Title: Structure and evolution of genes encoding polyubiquitin and ubiquitin-like p
A;Reference number: S55242; MUID:95229071
A;Accession: S55243
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-631 <CAL>
A;Cross-references: EMBL:L05917
A;Experimental source: ecotype Columbia
R;Callis, J.; Carpenter, T.; Sun, C.W.; Vierstra, R.D.
submitted to the EMBL Data Library, June 1995
A;Reference number: S61068
A;Accession: S61068
A;Molecule type: DNA
A;Residues: 1-341,'E',343-631 <CAW>
A;Cross-references: EMBL:L05917; NID:g870793; PIDN:AAA68879.1; PID:g870794
C;Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology
F;3-78/Domain: ubiquitin homology <UBH1>
F;79-154/Domain: ubiquitin homology <UBH2>
F;155-237/Domain: ubiquitin homology <UBH3>
F;238-318/Domain: ubiquitin homology <UBH4>
F;319-392/Domain: ubiquitin homology <UBH5>
F;393-468/Domain: ubiquitin homology <UBH6>
F;469-551/Domain: ubiquitin homology <UBH7>
F;552-627/Domain: ubiquitin homology <UBH8>

Query Match   20.3%;   Score 80;   DB 2;   Length 631;
Best Local Similarity   31.7%;   Pred. No. 0.32;
Matches   19;   Conservative   16;   Mismatches   25;   Indels   0;   Gaps   0;

```
Qy   11 GKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDYEIHDGMNLEL 70
          || :: ::|| ::|: :|:  ||  | : ::|||||||:|::|:||||
Db  561 GKTILEVESSDTIANVKEKIQVKEGIKPDQQMLIFFGQQLEDGVTLGDYIHKKSTLYL 620
```

RESULT   6
S60529
envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-14-13) (f
C;Species: human immunodeficiency virus type 1, HIV-1
A;Variety: isolate CI-14-13
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C;Accession: S60529
R;Janssen, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J
AIDS 8, 21-26, 1994
A;Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cot
A;Reference number: S60521; MUID:94280700
A;Accession: S60529
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-299 <JAN>
A;Cross-references: EMBL:X72031; NID:g468637; PIDN:CAA50914.1; PID:g468638
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Genetics:
A;Gene: env
C;Superfamily: type E retrovirus env polyprotein

C:Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match   17.9%;   Score 70.5; DB 2; Length 299;
Best Local Similarity 27.7%;   Pred. No. 1.7;
Matches 23; Conservative 11; Mismatches 20; Indels 29; Gaps 3;

Qy   9 RLGKKVRVKCN----------------TDDTIGDLKKLIAAQTGTRWNKIV----- 43
         :|||:||: | ||||||||||||||| : |||:
Db 20 QLVKPVRINCTRPSNNTRKSVPIGPGRAFYATDDIIGDIRQAHCNVSKTEWNETLQQVAT 79

Qy 44 -LKKWY---TIFKDHVSLGDYEI 62
         ||| | : ||||||:
Db 80 QLKKHFKNATIIFANPSGGDLEV 102

RESULT 7
Q0528
envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-14-5 and iso
C:Species: human immunodeficiency virus type 1, HIV-1
A:Variety: isolate CI-14-5; isolate CI-14-21
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C:Accession: S60528; S60530
R:Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J.;
AIDS 8, 21-26, 1994
A:Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cote d
A:Reference number: S60521; MUID:94280700
A:Accession: S60528
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-299 <JAN>
A:Cross-references: EMBL:X72030; NID:g468635; PIDN:CAA50913.1; PID:g468636
A:Experimental source: isolate CI-14-5
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
A:Accession: S60530
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-299 <JAW>
A:Cross-references: EMBL:X72032; NID:g468639; PIDN:CAA50915.1; PID:g468640
A:Experimental source: isolate CI-14-21
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C:Genetics:
A:Gene: env
C:Superfamily: type E retrovirus env polyprotein
C:Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match   17.9%;   Score 70.5; DB 2; Length 299;
Best Local Similarity 27.7%;   Pred. No. 1.7;
Matches 23; Conservative 11; Mismatches 20; Indels 29; Gaps 3;

Qy   9 RLGKKVRVKCN----------------TDDTIGDLKKLIAAQTGTRWNKIV----- 43
         :|||:||: | ||||||||||||||| : |||:
Db 20 QLVKPVRINCTRPSNNTRKSVPIGPGQAFYATDDIIGDIRQAHCNVSKTEWNETLRQVAT 79

Qy 44 -LKKWY---TIFKDHVSLGDYEI 62
         ||| | : ||||||:
Db 80 QLKKHFKNATIIFANPSGGDLEV 102

RESULT 8
T29404
hypothetical protein C16C8.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R:Waterston, R.; Le, T.T.; Gattung, S.
submitted to the EMBL Data Library, November 1996
A:Description: The sequence of C. elegans cosmid C16C8.
A:Reference number: Z20617
A:Accession: T29404
A:Molecule type: DNA
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA

A:Residues: 1-226 <WAT>
A:Cross-references: EMBL:U80452; PIDN:AAB37860.1; GSPDB:GN00020; CESP:C16C8.4
A:Experimental source: strain Bristol N2; clone C16C8
C:Genetics:
A:Gene: CESP:C16C8.4
A:Map position: 2
A:Introns: 27/3; 199/1

Query Match   17.5%;   Score 69; DB 2; Length 226;
Best Local Similarity 26.5%;   Pred. No. 1.9;
Matches 18; Conservative 13; Mismatches 29; Indels 8; Gaps 2;

Qy 10 LGKKVRVKCNTDTIGDLKKLIAAQ----TGTRWNKIVLKKWYTIFKDHVSLGDYEIHDG 65
         || |:||| :||| ||:||||| |||||| : | :|||:|||
Db 161 LGVSYAFKIHREDTVFDIKNDIEHRHDIPQHSYWLSFSGKR----LEDHCSIGDYNIQKS 216

Qy 66 MNLELYYQ 73
         : :|:
Db 217 STITMYFR 224

RESULT 9
A70363
mannose-1-phosphate guanyltransferase - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: A70363
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: A70363
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-831 <AQF>
A:Cross-references: GB:AE000704; NID:g2983301; PIDN:AAC06893.1; PID:g2983302; GB:AE00
A:Experimental source: strain VF5
C:Genetics:
A:Gene: mpg

Query Match   17.4%;   Score 68.5; DB 2; Length 831;
Best Local Similarity 27.4%;   Pred. No. 8.3;
Matches 23; Conservative 17; Mismatches 21; Indels 23; Gaps 5;

Qy   4 VVCND-RLGKKVRVK--------CNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDH 54
         ||: | |||||||| |:||||||||||||| |||||||||||||||||||:
Db 322 VICNDVKIGKRVKAKEGVVIAEDCEVEDEVLFLKDVVV------WPEKVIEKGSVVTKNI 375

Qy 55 VSLGDYE--IHDG-----MNLEL 70
         : |||: :|:| :|:||
Db 376 VCESKWEKGIFKGNKVIGRINVEL 399

RESULT 10
S60545
envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-45-1) (fr
C:Species: human immunodeficiency virus type 1, HIV-1
A:Variety: isolate CI-45-1
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C:Accession: S60545
R:Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J
AIDS 8, 21-26, 1994
A:Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cot
A:Reference number: S60545
A:Accession: S60545
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-294 <JAN>
A:Cross-references: EMBL:X72047; NID:g468669; PIDN:CAA50930.1; PID:g468670
A:Experimental source: isolate CI-45-1

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Genetics:
A;Gene: env
C;Superfamily: type E retrovirus env polyprotein
C;Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match   17.3%;   Score 68;   DB 2;   Length 294;
Best Local Similarity   26.8%;   Pred. No. 3.2;
Matches   22; Conservative   11; Mismatches   21; Indels   28; Gaps   3;

Qy   9 RLGKKVRVKCN--------------TDDFTIGDLKKLIAAQTGTRWNKIV----- 43
            :|:|::| :||:||:
Db   20 QLAKPVKINCTRPNNNTRKSVRIGPGQTFYATGDIIGDIRQAHCNVSGTEWNETLQKVAA 79

Qy   44 -LKKWY--TIFKDHVSLGDYEI 62
           |:|: :|||| ||
Db   80 QLRKHFNKTIIFASPSGGDVEI 101

RESULT 11
S60524
envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-45-3) (fragm
C;Species: human immunodeficiency virus type 1, HIV-1
A;Variety: isolate CI-45-3
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C;Accession: S60524
R;Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J.;
AIDS 8, 21-26, 1994
A;Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cote d
A;Reference number: S60521; MUID:94280700
A;Accession: S60524
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-294 <JAN>
A;Cross-references: EMBL:X72027; NID:g468780; PIDN:CAA50910.1; PID:g468781
A;Experimental source: isolate CI-45-3
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Genetics:
A;Gene: env
C;Superfamily: type E retrovirus env polyprotein
C;Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match   17.3%;   Score 68;   DB 2;   Length 294;
Best Local Similarity   26.8%;   Pred. No. 3.2;
Matches   22; Conservative   11; Mismatches   21; Indels   28; Gaps   3;

Qy   9 RLGKKVRVKCN--------------TDDFTIGDLKKLIAAQTGTRWNKIV----- 43
            :|:|::| :||:||:
Db   20 QLAKPVKINCTRPNNNTRKSVRIGPGQTFYATGDIIGDIRQARCNVSGTEWNETLQKVAA 79

Qy   44 -LKKWY--TIFKDHVSLGDYEI 62
           |:|: :|||| ||
Db   80 QLRKHFNKTIIFASPSGGDVEI 101

RESULT 12
S60550
envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-326-4) (frag
C;Species: human immunodeficiency virus type 1, HIV-1
A;Variety: isolate CI-326-4
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C;Accession: S60550
R;Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J.;
AIDS 8, 21-26, 1994
A;Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cote d
A;Reference number: S60521; MUID:94280700
A;Accession: S60550
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-303 <JAN>
A;Cross-references: EMBL:X72054; NID:g468681; PIDN:CAA50935.1; PID:g468682

A;Experimental source: isolate CI-326-4
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Genetics:
A;Gene: env
C;Superfamily: type E retrovirus env polyprotein
C;Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match   16.8%;   Score 66;   DB 2;   Length 303;
Best Local Similarity   26.4%;   Pred. No. 5.6;
Matches   23; Conservative   10; Mismatches   20; Indels   34; Gaps   4;

Qy   9 RLGKKVRVKC--------------NTDDTIGDLKKLIAAQTGTRWNKIV-- 43
            :|:|::| :||:||::
Db   20 QLSKPVRINCTRPGNNTRKSWRSWHIGPGRAFYTTDGIIGDIRKAYCNVSKTVWKDTLKQ 79

Qy   44 ----LKKWYT---IFKDHVSLGDYEI 62
             :|:|: :| |||| ||
Db   80 VARQLRKYFNNNTIIFANH-SGGDVEI 105

RESULT 13
S60549
envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-326-3) (f
C;Species: human immunodeficiency virus type 1, HIV-1
A;Variety: isolate CI-326-3
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C;Accession: S60549
R;Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J
AIDS 8, 21-26, 1994
A;Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cot
A;Reference number: S60521; MUID:94280700
A;Accession: S60549
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-303 <JAN>
A;Cross-references: EMBL:X72053; NID:g468679; PIDN:CAA50934.1; PID:g468680
A;Experimental source: isolate CI-326-3
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Genetics:
A;Gene: env
C;Superfamily: type E retrovirus env polyprotein
C;Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match   16.8%;   Score 66;   DB 2;   Length 303;
Best Local Similarity   26.4%;   Pred. No. 5.6;
Matches   23; Conservative   10; Mismatches   20; Indels   34; Gaps   4;

Qy   9 RLGKKVRVKC--------------NTDDTIGDLKKLIAAQTGTRWNKIV-- 43
            :|:|::| :||:||::
Db   20 QLSKPVRINCTRPGNNTRKSWRSWHIGPGRAFYTTDGIIGDIRKAYCNVSKTVWKDTLKQ 79

Qy   44 ----LKKWYT---IFKDHVSLGDYEI 62
             :|:|: :| |||| ||
Db   80 VARQLRKYFNNNTIIFANH-SGGDVEI 105

RESULT 14
T52254
cytochrome P450 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 01-Dec-2000
C;Accession: T52254
R;Maughan, J.A.; Nugent, J.H.A.; Hallahan, D.L.
submitted to the EMBL Data Library, May 1996
A;Description: Expression of CYP71B7 a cytochrome P450 expressed sequence tag from Ar
A;Reference number: Z26004
A;Accession: T52254
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-504 <MAU>
A;Cross-references: EMBL:X97864; PIDN:CAA66458.1

A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: CYP71B7
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: heme; iron; metalloprotein
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match   16.6%; Score 65.5; DB 2; Length 504;
Best Local Similarity   28.0%; Pred. No. 11;
Matches 21; Conservative 16; Mismatches 23; Indels 15; Gaps   3;

Qy   1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQ--TGTRWNKIVLKKWYTIF-------- 51
            | |::||::| | ::|||:: | :||:|||:||: | : |:|:|:|::: :
Db   319 MAELIRNPRVMKKVQDEIRT--TLGDKKQRITEQDLSQVHYFKLVVKEIFRLHPAAPLLL 376

Qy   52 ---KDHVSLGDYEI 62
             ||:|||:||:|
Db   377 PRETMSHVKIQGYDI 391

RESULT 15
D70581
probable cation-transporting ATPase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: D70581
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Skelton, S.; Squares, K.; Seeger, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: D70581
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-797 <COL>
A;Cross references: GB:Z95210; GB:AL123456; NID:g3261757; PIDN:CAB08506.1; PID:g2078057
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: ctpE
C;Superfamily: Na+/K+-transporting ATPase alpha chain; ATPase nucleotide-binding domain
F;436-601/Domain: ATPase nucleotide-binding domain homology <ATN>

Query Match   16.6%; Score 65.5; DB 2; Length 797;
Best Local Similarity   29.6%; Pred. No. 17;
Matches 24; Conservative 14; Mismatches 22; Indels 21; Gaps   5;

Qy   4 VVCNDRLG--KKVRVKCNTDD-TIGDLKKLIAAQTGTR--WNKIVLKKWYTIFKDHVS 56
           ::|:||::| ||::| | :|| | :||:|||::||| |:||| | :
Db   74 IIANSVIGMVQEIRAKQTLDKLAIIGQAKPLVRRQSGTRTRSTNEVVL-------DDIIE 126

Qy   57 LG-------DYEIHDGMNLEL 70
           || | |:|||:||:
Db   127 LGPGDQVVVDGEVVEEENLEI 147

Search completed: June 13, 2001, 20:27:07
Job time: 344 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2001, 18:35:07 ; Search time 1697.79 Seconds
                                         (without alignments)
                                         1759.785 Million cell updates/sec

Title: US-09-331-930A-1
Perfect score: 342
Sequence: 1 gttccaggagattacagctc........attttttgttaagtgtact 342

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

rched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*
               1: gb_est1:*
               2: gb_est2:*
               3: gb_est3:*
               4: gb_est4:*
               5: gb_est5:*
               6: gb_est6:*
               7: gb_est7:*
               8: gb_est8:*
               9: gb_est9:*
              10: gb_est10:*
              11: gb_est11:*
              12: gb_est12:*
              13: gb_est13:*
              14: gb_est14:*
              15: gb_est15:*
              16: gb_est16:*
              17: gb_est17:*
              18: gb_est18:*
              19: gb_est19:*
              20: gb_est20:*
              21: gb_est21:*
              22: gb_est22:*
              23: gb_est23:*
              24: gb_est24:*
              25: gb_est33:*
              26: gb_est34:*
              27: gb_est35:*
              28: gb_est36:*
              29: gb_est37:*
              30: gb_est38:*
              31: gb_est39:*
              32: gb_est40:*
              33: em_estba:*
              34: em_estfun:*
              35: em_esthum1:*
              36: em_esthum2:*
              37: em_esthum3:*
              38: em_esthum4:*
              39: em_esthum5:*
              40: em_esthum6:*
              41: em_esthum7:*
              42: em_esthum8:*
              43: em_esthum9:*

              44: em_esthum10:*
              45: em_esthum11:*
              46: em_esthum12:*
              47: em_esthum13:*
              48: em_esthum14:*
              49: em_esthum15:*
              50: em_esthum16:*
              51: em_esthum17:*
              52: em_esthum18:*
              53: em_esthum19:*
              54: em_esthum20:*
              55: em_esthum21:*
              56: em_esthum22:*
              57: em_esthum23:*
              58: em_esthum24:*
              59: em_esthum25:*
              60: em_esthum26:*
              61: em_esthum27:*
              62: em_esthum28:*
              63: em_estin1:*
              64: em_estin2:*
              65: em_estin3:*
              66: em_estin4:*
              67: em_estin5:*
              68: em_estom1:*
              69: em_estom2:*
              70: em_estov1:*
              71: em_estov2:*
              72: em_estpl1:*
              73: em_estpl2:*
              74: em_estpl3:*
              75: em_estpl4:*
              76: em_estpl5:*
              77: em_estpl6:*
              78: em_estpl7:*
              79: em_estpl8:*
              80: em_estpl9:*
              81: em_estpl10:*
              82: em_estro1:*
              83: em_estro2:*
              84: em_estro3:*
              85: em_estro4:*
              86: em_estro5:*
              87: em_estro6:*
              88: em_estro7:*
              89: em_estro8:*
              90: em_estro9:*
              91: em_estro10:*
              92: em_estro11:*
              93: em_estro12:*
              94: em_estro13:*
              95: em_estro14:*
              96: em_estro15:*
              97: em_estro16:*
              98: em_estro17:*
              99: em_estro18:*
             100: em_estro19:*
             101: em_estro20:*
             102: gb_est25:*
             103: gb_est26:*
             104: gb_est27:*
             105: gb_est28:*
             106: gb_est29:*
             107: gb_est30:*
             108: gb_est31:*
             109: gb_est32:*
             110: gb_est41:*
             111: gb_est42:*
             112: gb_est43:*
             113: gb_est44:*
             114: gb_est45:*
             115: gb_est46:*
             116: gb_est47:*

```
117: gb_est48:*           190: em_gss_pln1:*
118: gb_est49:*           191: em_gss_pln2:*
119: gb_est50:*           192: em_gss_pro:*
120: gb_est51:*           193: em_gss_rod1:*
121: gb_est52:*           194: em_gss_rod2:*
122: gb_est53:*           195: em_gss_rod3:*
123: gb_est54:*           196: em_gss_rod4:*
124: gb_est55:*           197: em_gss_rod5:*
125: gb_est56:*           198: em_gss_vrt1:*
126: gb_est57:*           199: em_gss_vrt2:*
127: gb_est58:*           200: em_gss_vrt3:*
128: gb_est59:*           201: gb_gss1:*
129: gb_est60:*           202: gb_gss2:*
130: gb_est61:*           203: gb_gss3:*
131: gb_est62:*           204: gb_gss4:*
132: gb_est63:*           205: gb_gss5:*
133: gb_est64:*           206: gb_gss6:*
134: gb_est65:*           207: gb_gss7:*
135: gb_est66:*           208: gb_gss8:*
136: gb_est75:*           209: gb_gss9:*
137: gb_est76:*           210: gb_gss10:*
138: gb_est77:*           211: gb_gss11:*
139: gb_est78:*           212: gb_gss12:*
140: gb_est79:*           213: gb_gss13:*
141: gb_est80:*           214: gb_gss14:*
142: gb_est81:*           215: gb_gss15:*
143: gb_est82:*           216: gb_gss16:*
144: gb_est83:*           217: gb_gss17:*
145: gb_est84:*           218: gb_gss18:*
146: gb_est85:*           219: gb_gss19:*
147: gb_est86:*           220: gb_gss20:*
148: gb_est87:*           221: gb_gss21:*
149: gb_est88:*           222: gb_gss22:*
150: gb_est89:*           223: gb_gss23:*
151: gb_est90:*           224: gb_gss24:*
152: gb_est99:*           225: gb_gss25:*
153: gb_est100:*          226: gb_gss26:*
154: gb_est101:*          227: gb_gss27:*
155: gb_est102:*          228: gb_gss28:*
156: gb_est103:*          229: gb_gss29:*
157: gb_est104:*          230: gb_gss30:*
158: gb_est105:*          231: gb_gss31:*
159: gb_est106:*          232: gb_gss32:*
160: gb_est67:*           233: gb_gss33:*
161: gb_est68:*           234: gb_gss34:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
```

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.		Score	% Query Match	Length	DB	ID	Description
1		254.8	74.5	382	4	AA241830	AA241830 mx12b09.r
2	c	254.8	74.5	384	137	BE626037	BE626037 ut81c03.x
3		254.8	74.5	392	4	AA265956	AA265956 mz72e08.r
4		254.8	74.5	392	4	AA275337	AA275337 vc07d10.r
5	c	254.8	74.5	394	20	AI463272	AI463272 uc44f05.x
6	c	254.8	74.5	395	142	BF017894	BF017894 uw84e06.x
7		254.8	74.5	397	158	W17970	W17970 mb82b05.r1
8	c	254.8	74.5	398	137	BE628007	BE628007 uu26g05.x
9		254.8	74.5	398	159	W85207	W85207 mf50g10.r1
10	c	254.8	74.5	398	164	BE197360	BE197360 ug69c04.x
11		254.8	74.5	399	167	BE448243	BE448243 ut81c03.y
12		254.8	74.5	403	1	AA050441	AA050441 mj12d12.r
13		254.8	74.5	406	7	AA462867	AA462867 vf91h07.r
14		254.8	74.5	410	158	W83387	W83387 mf28c03.r1
15		254.8	74.5	411	158	W11593	W11593 ma90a06.r1
16		254.8	74.5	421	158	W11035	W11035 ma76d01.r1
17		254.8	74.5	421	158	W57187	W57187 md67g03.r1
18		254.8	74.5	422	7	AA435358	AA435358 ve15e05.r

```
        19  254.8  74.5  423  12  AA792859          AA792859 vs89h01.r
        20  254.8  74.5  425   1  AA007837          AA007837 mg63g09.r
        21  254.8  74.5  426   1  AA060970          AA060970 mj86e07.r
        22  254.8  74.5  430  158  W13465           W13465 mb34h10.r1
        23  254.8  74.5  432   7  AA461775          AA461775 vf94c08.r
        24  254.8  74.5  434  158  W81897           W81897 me92b03.r1
        25  254.8  74.5  436  158  W41537           W41537 mc45g01.r1
        26  254.8  74.5  440  14  AI005986          AI005986 ua84e08.r
        27  254.8  74.5  441   1  AA028841          AA028841 mh87h10.r
        28  254.8  74.5  441  158  W70638           W70638 me34c10.r1
     c  29  254.8  74.5  451  172  BG066897         BG066897 H3048C09-
        30  254.8  74.5  471   4  AA271476          AA271476 va83b12.r
        31  254.8  74.5  482  173  BG080039         BG080039 H3048C09-
        32  254.3  74.3  422  158  W29753           W29753 mc07d08.r1
        33  254.2  74.3  434  158  W41069           W41069 mc39b05.r1
        34  254    74.3  415  137  BE624821         BE624821 uu26g05.y
     c  35  253.8  74.2  432  175  C80874           C80874 C80874 Mous
        36  253.2  74.0  399  158  W53910           W53910 md12e08.r1
        37  253.2  74.0  410  137  BE625559         BE625559 uu27a10.y
        38  253.2  74.0  422  117  AW611154         AW611154 un77h05.y
        39  251.8  73.6  341  159  W97247           W97247 mg07b11.r1
        40  251.8  73.6  394   8  AA512172          AA512172 vj21b02.r
        41  251.6  73.6  423  11  AA727805          AA727805 vp33a08.r
        42  250.6  73.3  384  16  AI121718          AI121718 uc44f05.r
        43  248    72.5  412   9  AA623135          AA623135 vi13a12.r
        44  246.2  72.0  338  165  BE334800         BE334800 us89d03.y
     c  45  245    71.6  393  148  BF456049         BF456049 UI-M-B21-
```

ALIGNMENTS

```
RESULT   1
AA241830
LOCUS       AA241830      382 bp      mRNA                    EST       07-MAR-1997
DEFINITION  mx12b09.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:679961
            5', mRNA sequence.
ACCESSION   AA241830
VERSION     AA241830.1  GI:1872932
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 382)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Schellenberg,K., Tan,F., Lacy,M., Le,M., Martin,J., Morris,M.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:419665
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..382
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:679961"
                     /clone_lib="Soares mouse NML"
                     /tissue_type="liver"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
```

was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGGCCGCGGGCCGGAATCTTTTTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

```
BASE COUNT     96 a     95 c     91 g    100 t
ORIGIN

Query Match                    74.5%;   Score 254.8; DB 4;  Length 382;
Best Local Similarity   89.3%;  Pred. No. 4.2e-68;
Matches  300;  Conservative   0;  Mismatches  27;  Indels   9;  Gaps   2;

Qy    1  gttcaggagattacagctccagccacacaatgattgaggtggttgcaacgaccgtctagg  60
         |||| |||  ||| |||||||||||||||||| ||||| ||||||||||||||||||| ||
Db    9  GTTCGGAGATATTGGAGCTCCAGCCACAATGATTGAGGTGGTTGCAACGACCGTCTCGG  68

Qy   61  aaagaaagtccgcgttaagtgcaacacaccgatgacaccatcgggagacttgaagaaactgat  120
         ||||||||| |||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   69  AAAGAAAGTCCGCGTTAAGTGCAACACCGATGACACCATCGGCGACTTGAAGAAACTGAT  128

Qy  121  agcggccaaactggcactcgttggaataagatcgttcttaaaagtgtacgatttt  180
         ||| ||  ||||||||||| ||  ||||||||||||||||||||||||||||||||
Db  129  AGCTGCTCAAACTGGCACCCGTGGACAACAAGATCGTTCTTAAAAAGTGGTACACGATTTT  188

Qy  181  taaggaccatgtatctctgggagattatgaaatccacgatggatgaactggagctta  240
         |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db  189  TAAGGACCACGTGTCTCTGGGAGATTATGAAATCCACGATGGATGAACCTGGAGCTTA  248

Qy  241  ttaccagtagagggaattcctccacct------tgccaaccttgctttcctctccc  292
         |||||||||||| |||| |||  |   ||     |  |  |||||||||||||||||
Db  249  TTACCAGTAGAGGGGGATTCCTTCCTCCTCCGCCCTGCTCTGCCCTGCCCTCCTCTCCC  308

Qy  293  atggctcattaacactgttgtagatgctcatttt  328
         || |||||| |||||||  |||||| |||||||||
Db  309  AT-CCTCATCTGACACTGGTGTAGATGGTCATTTTT  343
```

```
RESULT   2
BE626037/c
LOCUS       BE626037       384 bp      mRNA                              24-AUG-2000
DEFINITION  ut81c03.x1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:3368836 3' similar to WP:F46F11.4 CE10602 ;, mRNA sequence.
ACCESSION   BE626037
VERSION     BE626037.1  GI:9905453
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 384)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: ut81c03.y1
            Contact: Robert Strausberg, Ph.D.
            Email: Robert_Strausberg@nih.gov
            Tel: (301) 496-1550
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1078440
FEATURES             Location/Qualifiers
     source          1..384
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3368836"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
```

FEATURES   Location/Qualifiers
 source   1..392
     /organism="Mus musculus"
     /strain="C57BL/6J"
     /db_xref="taxon:10090"
     /clone="IMAGE:718982"
     /clone_lib="Soares mouse lymph node NbMLN"
     /sex="male"
     /tissue_type="lymph node"
     /dev_stage="4 weeks"
     /lab_host="DH10B"
     /note="Organ: lymph node; Vector: pT7T3D-Pac (Pharmacia)
     with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
     1st strand cDNA was primed with a Not I - oligo(dT) primer
     [5'
     TGTTACCAATCTGAAGTGGGAGCGGCGCCGCGATACTTTTTTTTTTTTTTTTTTTT
     3']; double-stranded cDNA was ligated to Eco RI adaptors
     (Pharmacia), digested with Not I and cloned into the Not I
     and Eco RI sites of the modified pT7T3 vector. RNA
     provided by Dr. Bertrand Jordan. Library constructed and
     normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT   91 a   102 c   99 g   100 t
ORIGIN

Query Match   74.5%;   Score 254.8;   DB 4;   Length 392;
Best Local Similarity   89.3%;   Pred. No. 4.2e-68;
Matches 300;   Conservative   0;   Mismatches   27;   Indels   9;   Gaps   2;

```
Qy     1 gttccaggagattacagctccagccacaatgattgaggtggttgcaacgaccgtctagg 60
           |||| ||| |||  |||||||||||||||||||||||||||||||||||||||||||||
Db    55 GTTCGAGGAGATATTGGAGCTCCAGCCACAATGATTGAGGTGGTTTGCAACGACCGTCTCGG 114

Qy    61 aagaaagtccgcgttaagtgcaacaccgatgacaccatcgggagacttgaagaaactgat 120
           |||||||||||||||||||||||||| ||||||||||||||| |||||||||||||||||
Db   115 AAGAAAGTCCGCGTTAAGTGCAACAACACCGATGACACCATCGGGACACTTGAAGAAACTGAT 174

Qy   121 agcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatttt 180
           ||||| |||||||||| || ||||||| | ||||||||||||||||| ||||||||||||
Db   175 AGCTGCTCAAACTGGCACCCGCTGGAACAAGATCGTTCTTAAAAAGTGTACACGATTTT 234

Qy   181 taaggaccatgtatctctggggagattatgaaatccacgatgggatgaacctggagcttta 240
           |||||||| || |||||  |||||||||| ||||||||||||||||||||||||||||||
Db   235 TAAGGACCACCGTGTCTCTCTGGGAGATTATGAAATCCACGATGGGATGAACCTGGAGCTTTA 294

Qy   241 ttaccagtagagggggaattcctccacct-------tgcccaacctttgctttcctctccc 292
           |||||||||||| |||||| ||| ||||||       |  |||||  ||| |||||||||
Db   295 TTACCAGTAGAGGGGGATTCCTTCTCCTCCTCGCCCTCTGCCCTGCCCTCCTCCTCCC 354

Qy   293 atggctcattaacactgttgtagatgctcattttt 328
           || ||||||||||| |||||||||||||||||||
Db   355 AT-CCTCATCTGACACTGGTGTAGAGATGGTCATTTTT 389
```

RESULT 4
AA275337
LOCUS   AA275337   392 bp   mRNA   EST   31-MAR-1997
DEFINITION   vc07d10.r1 Soares mouse lymph node NbMLN Mus musculus cDNA clone
     IMAGE:765811 5', mRNA sequence.
ACCESSION   AA275337
VERSION   AA275337.1   GI:1915823
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1   (bases 1 to 392)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
     Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
     Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
     Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
     Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project

     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
     polylinker; 1st strand cDNA was prepared from mammary
     gland tissue from a lactating female, and was then primed
     with a Not I - oligo(dT) primer. Double-stranded cDNA was
     ligated to Eco RI adaptors (Pharmacia), digested with Not
     I and cloned into the Not I and Eco RI sites of the
     modified pT7T3 vector. Library is normalized. Library
     was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT   98 a   90 c   95 g   101 t
ORIGIN

Query Match   74.5%;   Score 254.8;   DB 137;   Length 384;
Best Local Similarity   89.3%;   Pred. No. 4.2e-68;
Matches 300;   Conservative   0;   Mismatches   27;   Indels   9;   Gaps   2;

```
Qy     1 gttccaggagattacagctccagccacaatgattgaggtggttgcaacgaccgtctagg 60
           |||| ||| |||  |||||||||||||||||||||||||||||||||||||||||||||
Db   375 GTTCGAGGAGATATTGGAGCTCCAGCCACAATGATTGAGGTGGTTTGCAACGACCGTCTCGG 316

Qy    61 aagaaagtccgcgttaagtgcaacaccgatgacaccatcgggagacttgaagaaactgat 120
           |||||||||||||||||||||||||| ||||||||||||||| |||||||||||||||||
Db   315 AAGAAAGTCCGCGTTAAGTGCAACAACACCGATGACACCATCGGGACACTTGAAGAAACTGAT 256

Qy   121 agcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatttt 180
           ||||| |||||||||| || ||||||| | ||||||||||||||||| ||||||||||||
Db   255 AGCTGCTCAAACTGGCACCCGCTGGAACAAGATCGTTCTTAAAAAGTGTACACGATTTT 196

Qy   181 taaggaccatgtatctctggggagattatgaaatccacgatgggatgaacctggagcttta 240
           |||||||| || |||||  |||||||||| ||||||||||||||||||||||||||||||
Db   195 TAAGGACCACCGTGTCTCTCTGGGAGATTATGAAATCCACGATGGGATGAACCTGGAGCTTTA 136

Qy   241 ttaccagtagagggggaattcctccacct-------tgcccaacctttgctttcctctccc 292
           |||||||||||| |||||| ||| ||||||       |  |||||  ||| |||||||||
Db   135 TTACCAGTAGAGGGGGATTCCTTCTCCTCCTCGCCCTCTGCCCTGCCCTCCTCCTCCC 76

Qy   293 atggctcattaacactgttgtagatgctcattttt 328
           || ||||||||||| |||||||||||||||||||
Db    75 AT-CCTCATCTGACACTGGTGTAGAGATGGTCATTTTT 41
```

RESULT 3
AA265956
LOCUS   AA265956   392 bp   mRNA   EST   20-MAR-1997
DEFINITION   mz72e08.r1 Soares mouse lymph node NbMLN Mus musculus cDNA clone
     IMAGE:718982 5', mRNA sequence.
ACCESSION   AA265956
VERSION   AA265956.1   GI:1902046
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1   (bases 1 to 392)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
     Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
     Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
     Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
     Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   Contact: Marra M/Mouse EST Project
     WashU-HHMI Mouse EST Project
     Washington University School of MedicineP
     4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
     Tel: 314 286 1800
     Fax: 314 286 1810
     Email: mouseest@watson.wustl.edu
     This clone is available royalty-free through LLNL; contact the
     IMAGE Consortium (info@image.llnl.gov) for further information.
     MGI:444478
     Seq primer: -28m13 rev2 ET from Amersham
     High quality sequence stop: 326.

JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:466731
 Seq primer: -28m13 rev2 ET from Amersham.

FEATURES Location/Qualifiers
 source 1..392
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:765811"
 /clone_lib="Soares mouse lymph node NbMLN"
 /sex="male"
 /tissue_type="lymph node"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Organ: lymph node; Vector: pT7T3D-Pac (Pharmacia)
 with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 1st strand cDNA was primed with a Not I - oligo(dT) primer
 [5'
 TGTTACCAATCTGAAGTGGGAGGCGGCCGCGATACTTTTTTTTTTTTTTTTTTTTTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. RNA
 provided by Dr. Bertrand Jordan. Library constructed and
 normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 92 a 102 c 98 g 100 t
ORIGIN

Query Match 74.5%; Score 254.8; DB 4; Length 392;
Best Local Similarity 89.3%; Pred. No. 4.2e-68;
Matches 300; Conservative 0; Mismatches 27; Indels 9; Gaps 2;

```
Qy    1 gttccaggagattacagctccagccacaatgattgaggtggttgcaacgaccgtctcagg 60
        |||| ||| ||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   54 GTTCGAGGATATTGGAGCTCCAGCCACAATGATTGAGGTGGTTGCAACGACCGTCTCGG 113

Qy   61 aaagaaagtccgcgttaagtgcaacaccgatgacaccggggacttgaagaaactgat 120
        ||||||||||||||||||||| |||||||||||||| |||||||||||||||||||
Db  114 AAAGAAAGTCCGCGTTAAGTGCAACACCGATGACACCGATAAGGTGGTACACGATGAT 173

Qy  121 agcggccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgattt 180
        |||| || |||||||||  ||||||||||||||||||||||||||||||||||||
Db  174 AGCTGCTCAAACTGGCACCCGCTGGAACAAGATCGTTCTTAAAAAGTGGTACACGATTTT 233

Qy  181 taaggaccatgtatctctggagattatgaaatccacgatggatgaacctggagcttta 240
        ||||||| | |||||||||||||||||||||||||||||||||||||||||||||
Db  234 TAAGGACCACCGTGTCTCTGGAGATTATGAAATCCACGATGGATGAACCTGGAGCTTTA 293

Qy  241 ttaccagtagaggggaattcctccacct------tgcccaaccttgctttcctcccc 292
        |||||||||||||| ||||| ||  || ||  |||||||||||||||||||||||
Db  294 TTACCAGTAGAGGGGGATTCCTTCCTTCCTCGCCCCTGCTCTCGCCCGCCCTCCTCCCC 353

Qy  293 atggctcatttaacactgttgtagatgctcattttt 328
        ||  ||||| || |||||||||||||||||||||||
Db  354 AT-CCTCATCTGACACTGGTGTAGATGCTCATTTTT 388
```

RESULT 5
AI463272/c
LOCUS AI463272 394 bp mRNA EST 09-MAR-1999
DEFINITION uc44f05.x1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
 IMAGE:1400865 3' similar to WP:F46F11.4 CE10602 ;, mRNA sequence.
ACCESSION AI463272
VERSION AI463272.1 GI:4317302

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 394)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
 Tel.: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:912581
 This clone was previously sequenced on the 5' end only, this new
 data is from the 3' end
 High quality sequence stop: 376.

FEATURES Location/Qualifiers
 source 1..394
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:1400865"
 /clone_lib="Soares_mammary_gland_NMLMG"
 /sex="female (lactating)"
 /tissue_type="mammary gland"
 /lab_host="DH10B"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; 1st strand cDNA was prepared from mammary
 gland tissue from a lactating female, and was then primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT7T3 vector. Library is normalized. Library
 was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 101 a 95 c 96 g 102 t
ORIGIN

Query Match 74.5%; Score 254.8; DB 20; Length 394;
Best Local Similarity 89.3%; Pred. No. 4.2e-68;
Matches 300; Conservative 0; Mismatches 27; Indels 9; Gaps 2;

```
Qy    1 gttccaggagattacagctccagccacaatgattgaggtggttgcaacgaccgtctagg 60
        |||| ||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db  370 GTTCGAGGATATTGGAGCTCCAGCCACAATGATTGAGGTGGTTGCAACGACCGTCTCGG 311

Qy   61 aaagaaagtccgcgttaagtgcaacaccgatgacacccgggacttgaagaaactgat 120
        ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  310 AAAGAAAGTCCGCGTTAAGTGCAACACCGATGACACCATCCGGGACTTGAAGAAACTGAT 251

Qy  121 agcggccaaactgcactcgttggaataagatcgttcttaaaaagtggtacacgatttt 180
        |||| || |||| ||||  ||||||||||||||||||||||||||||||||||||||
Db  250 AGCTGCTCAAACTGGCACCCCTGGAACAAGATCGTTCTTAAAAAGTGGTACACGATTTT 191

Qy  181 taaggaccatgtatctctggagattatgaaatccacgatggatggaacctggagctta 240
        ||||||| | |||||||||||||||||||||||||||||||||||| |||||||||||
Db  190 TAAGGACCACCGTGTCTCTGGAGATTATGAAATCCACGATGGATGAACCTGGAGCTTTA 131

Qy  241 ttaccagtagaggggaattcctccacct------tgcccaaccttgctttcctcccc 292
        ||||||||||||| ||||| ||  ||| ||  |||||||||||||||||||||||||
Db  130 TTACCAGTAGAGGGGGATTCCTTCCTTCCTCGCCCCTGCTCTCGCCCGCCCTCCTCCCC 71

Qy  293 atggctcattaacactgttgtagatgctcattttt 328
        ||  ||||| || |||||||||||||| |||||||
Db   70 AT-CCTCATCTGACACTGGTGTAGATGGTCATTTTT 36
```

RESULT 6
BF017894/c
LOCUS BF017894 395 bp mRNA EST 10-OCT-2000
DEFINITION uw84e06.x1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone

Left record

```
DEFINITION   IMAGE:3468706 3' similar to TR:Q9V998 Q9V998 CG3450 PROTEIN. ;
             mRNA sequence.
ACCESSION    BF017894
VERSION      BF017894.1  GI:10749226
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 395).
  AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: Robert_Strausberg@nih.gov
             Tel: (301) 496-1550
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:1388066.
FEATURES             Location/Qualifiers
     source          1..395
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3468706"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library is normalized. Library
                     was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      102 a     96 c     97 g    100 t
ORIGIN

Query Match           74.5%;  Score 254.8;  DB 142;  Length 395;
Best Local Similarity 89.3%;  Pred. No. 4.2e-68;
Matches  300;  Conservative   0;  Mismatches  27;  Indels   9;  Gaps   2;

Qy      1 gttccaggagattacagctccagccacaatgattgagtggtttgcaacgaccgtctagg 60
          ||||| ||| ||| ||||||||||||||||||||||||||| ||||||||||||||| ||
Db    368 GTTCGAGGATATTGGAGCTCCAGCCACCACAATGATTGAGGTGGTTTGCAACGACCGTCTCGG 309

Qy     61 aaagaaagtccgcgttaagtgcaacaccgatgacacaccgggacttgaagaaactgat 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    308 AAAGAAAGTCCGCGTTAAGTGCAACACCGATGACACACCGGGACTTGAAGAAACTGAT 249

Qy    121 agcggcccaaactggcactcgttggaataagatcgttccttaaaaagtggtacacgatttt 180
          || ||||| |||||||||| || ||||| |||||||| ||||||||||||||||||||||
Db    248 AGCTGCTCAAACTGGCACCCGCTGGAACAAGATCGTTCTTAAAAAGTGGTACACGATTTT 189

Qy    181 taaggaccatgtatctctgggagattgaaatccacgatgggatgaacctggagcttta 240
          |||||||| | ||| |||||||||| || ||||||||||||||||||||||||||||||
Db    188 TAAGGACCACCGTGTCTCTGGGAGATTATGAAAATCACGATGGGATGAACCTGGAGCTTTA 129

Qy    241 ttaccagtagagggaattcctccacct--------tgcccaacctgctttcctctccc 292
          |||||||||| |||||| |||| ||||        || || |||||| |||| ||||||
Db    128 TTACCAGTAGAGGGGGATTCCTTCCTCTCCTGCCCTGCCCTGGCCTGCCCTCCTCTCCC 69

Qy    293 atggtcatttaacactgttgtagatgctcattttt 328
          || |||| | ||||| || |||||||| ||||||
Db     68 AT-CCTCATCTGCACACTGGTGTAGATGGTCATTTT 34

RESULT  7
W17970
LOCUS       W17970                   397 bp    mRNA            EST       12-SEP-1996
```

Right record

```
DEFINITION   mb82b05.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
             IMAGE:335889 5', mRNA sequence.
ACCESSION    W17970
VERSION      W17970.1  GI:1292354
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 397)
  AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R., and
             Waterston,R.
  TITLE      The WashU-HHMI Mouse EST Project
  JOURNAL    Unpublished (1996)
COMMENT      Contact: Marra M/Mouse EST Project
             WashU-HHMI Mouse EST Project
             Washington University School of MedicineP
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:21289
             Putative full length read
             vector to vector length is 398
             Seq primer: -28M13 rev2 from Amersham.
FEATURES             Location/Qualifiers
     source          1..397
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:335889"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site 1: Not I: Site 2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer (5'
                     TGTTACCAATCTGAAGTGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
BASE COUNT       99 a     99 c     97 g    102 t
ORIGIN

Query Match           74.5%;  Score 254.8;  DB 158;  Length 397;
Best Local Similarity 89.3%;  Pred. No. 4.3e-68;
Matches  300;  Conservative   0;  Mismatches  27;  Indels   9;  Gaps   2;

Qy      1 gttccaggagattacagctccagccacaatgattgagtggtttgcaacgaccgtctagg 60
          ||||| ||| ||| ||||||||||||||||||||||||||| ||||||||||||||| ||
Db     30 GTTCGAGGATATTGGAGCTCCAGCCACCACAATGATTGAGGTGGTTTGCAACGACCGTCTCGG 89

Qy     61 aaagaaagtccgcgttaagtgcaacaccgatgacacaccgggacttgaagaaactgat 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90 AAAGAAAGTCCGCGTTAAGTGCAACACCGATGACACACCGGGACTTGAAGAAACTGAT 149

Qy    121 agcggcccaaactggcactcgttggaataagatcgttccttaaaaagtggtacacgatttt 180
          || ||||| |||||||||| || ||||| |||||||| ||||||||||||||||||||||
Db    150 AGCTGCTCAAACTGGCACCCGCTGGAACAAGATCGTTCTTAAAAAGTGGTACACGATTTT 209

Qy    181 taaggaccatgtatctctgggagattgaaatccacgatgggatgaacctggagcttta 240
          |||||||| | ||| |||||||||| || ||||||||||||||||||||||||||||||
Db    210 TAAGGACCACCGTGTCTCTGGGAGATTATGAAAATCACGATGGGATGAACCTGGAGCTTTA 269
```

```
Qy    241 ttaccagtagagggaattcctccacct--------tgccaacctgcttcctcccc 292
          ||||||||||| ||||| ||| ||  |||||||||||||||||
Db    270 TTACCAGTAGAGGGGGATTCCTTCTCCTCCGCCCCTGCGCCCCTCGCCC 329

Qy    293 atggctcatttaacactgttgtagatgctcatttt 328
          || |||||| ||| ||||| ||||||||||||||||
Db    330 AT-CCTCATCTGACACTGGTGTAGATGGTCATTTT 364
```

RESULT 8
BE628007/c
```
LOCUS       BE628007      398 bp    mRNA      EST      24-AUG-2000
DEFINITION  uu26g05.x1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:3373112 3' similar to WP:F46F11.4 CE10602 ;, mRNA sequence.
ACCESSION   BE628007
VERSION     BE628007.1  GI:9908452
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 398)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1082716.
FEATURES    Location/Qualifiers
source      1..398
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="IMAGE:3373112"
            /clone_lib="Soares_mammary_gland_NMLMG"
            /sex="female (lactating)"
            /tissue_type="mammary gland"
            /lab_host="DH10B"
            /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
            polylinker; 1st strand cDNA was prepared from mammary
            gland tissue from a lactating female, and was then primed
            with a Not I - oligo(dT) primer. Double-stranded cDNA was
            ligated to Eco RI adaptors (Pharmacia), digested with Not.
            I and cloned into the Not I and Eco RI sites of the
            modified pT7T3 vector. Library is normalized. Library
            was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT    102 a    96 c    99 g    101 t
ORIGIN

Query Match       74.5%;   Score 254.8;   DB 137;  Length 398;
Best Local Similarity  89.3%;   Pred. No. 4.3e-68;
Matches 300;  Conservative  0;  Mismatches 27;  Indels  9;  Gaps  2;

Qy     1 gttccaggagattacagctccagccacaatgattgaggtgggttgcaacgacgtctagg 60
         |||| ||| |||| ||| |||||||||||||||||||||||||||||| ||
Db   370 GTTCGAGGAGATATTGGAGCTCCAGCCACCAAATGATTGAGGTGGTTGCAACGACCGTCCGG 311

Qy    61 aagaaagtccgcgttaagtgcaacaccgatgacaccatcgggacctgagaaaactgat 120
         ||||| |||||||||||| ||  | ||||||| || ||||||| ||||||||||||||
Db   310 AAAGAAGTCCGCGTTAAGTGCAACACCGATGACACCATCCGCGACACTGAAGAAACTGAT 251

Qy   121 agcggccaaactggcactgttggaataagatcgttcttaaaaagtggtacacgatttt 180
         || |||||||||||||| | |||||||| |||| |||||||| |||||||||  |||||
Db   250 AGCTGCTCAAACTGGCACCTGGCACCCCGCTGAACAGATCGTTCTTAAAAAGTGGTACACGATTTT 191

Qy   181 taaggaccatgtatctctgggagattatgaaatccacgatgggatgaacctggagcttta 240
         ||||||| || |||||||||||||| || || ||| |||| |||||||| |||||||||
Db   190 TAAGGACCACCGTGTCTCTGGGAGATTATGAAATCCACGATGGGATGGATGAACCTGGAGCTTTA 131
```

```
Qy    241 ttaccagtagagggaattcctccacct--------tgccaacctgcttcctcccc 292
          ||||||||||| ||||| ||| || |||||||||||||
Db    130 TTACCAGTAGAGGGGGATTCCTTCTCCTCCGCCCCTGCGCCCCTCGCCC 71

Qy    293 atggctcatttaacactgttgtagatgctcatttt 328
          || |||||| ||| ||||| ||||||||||||||||
Db     70 AT-CCTCATCTGACACTGGTGTAGATGGTCATTTT 36
```

RESULT 9
W85207
```
LOCUS       W85207      398 bp    mRNA      EST      12-SEP-1996
DEFINITION  mf50g10.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:408546 5', mRNA sequence.
ACCESSION   W85207
VERSION     W85207.1  GI:1541533
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 398)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE       The WashU-HHMI Mouse EST Project
JOURNAL     Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1397677.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:252314
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 392.
FEATURES    Location/Qualifiers
source      1..398
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="IMAGE:408546"
            /clone_lib="Soares mouse embryo NbME13.5 14.5"
            /sex="unknown"
            /tissue_type="embryo"
            /dev_stage="13.5-14.5dpc total fetus"
            /lab_host="DH10B"
            /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
            was primed with a Not I - oligo(dT) primer (5'
            TGTTACCAATCTGAAGTGGCAGCGGCCGCGGGAAATTTTTTTTTTTTTTTTTTTT
            T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
            14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
            State Univ., from 2 ]; double-stranded cDNA was ligated to
            Eco RI adaptors (Pharmacia), digested with Not I and
            cloned into the Not I and Eco RI sites of the modified
            pT7T3 vector. Library went through one round of
            normalization, and was constructed by Bento Soares and
            M.Fatima Bonaldo.
            "
BASE COUNT   100 a    99 c    97 g    102 t
ORIGIN

Query Match      74.5%;   Score 254.8;   DB 159;  Length 398;
Best Local Similarity  89.3%;   Pred. No. 4.3e-68;
Matches 300;  Conservative  0;  Mismatches 27;  Indels  9;  Gaps  2;
```

```
Qy    1 gttccaggagattacagctccagccacaatgattgaggtggtttgcaacgaccgtctagg 60
        ||||| ||| |||| |||| |||||||||||||||||||||||||||||||||| || ||
Db  370 GTTCGGAGGATATTGGAGCTCCAGCCACAATGATTGAGGTGGTTTGCAACGACCGTCTCGG 311

Qy   61 aagaaagtccgcgttaagtgcaacacaccgatgacaccatcgggacttgaagaaactgat 120
        |||||||||||| ||||||||||||||||||||||||||||| ||||||||||||||| ||
Db  310 AAAGAAAGTCCGGTTAAGTGCAACACCGATGACACCATCGGCGACTTGAAGAAACTGAT 251

Qy  121 agcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatttt 180
        || || ||||||||||||| |||||| |||||||||||||||||||||||||||||||||
Db  250 AGCTGCTCAAACTGGCACCCGCTGGAACAAGATCGTTCTTAAAAAGTGTACACGATTTT 191

Qy  181 taaggaccatgtatctctgggagattagaaatccacgatgggatgaacctggagctta 240
        |||||||||| || ||||||||||||| |||||||||||||||||||||||||||| ||
Db  190 TAAGGACCACCGTGTCTCTGGGAGATTATGAAATCCACGATGGGATGAACCTGGAGCTTTA 131

Qy  241 ttaccagtagagggggaattcctccacct------tgcccaacctgcttcctctctccc 292
        ||||||||||||||| |||||| |||||        ||||| ||||| |||||||||||
Db  130 TTACCAGTAGAGGGGGGATTCCTTCCTCCGCCCTGCTCTGCCCCTGCCCCTCTCCC 71

Qy  293 atggctcatttaacactgttgtagatgctcattttt 328
        || |||||| |||||||||||||||||||||||||||
Db   70 AT-CCTCATCTGCACACTGGTGTAGATGCTCATTTTT 36

RESULT 11
BE448243
LOCUS        BE448243        399 bp    mRNA                 EST      25-JUL-2000
DEFINITION   ut81c03.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
             IMAGE:3368836 5' similar to WP:F46F11.4 CE10602 ;, mRNA sequence.
ACCESSION    BE448243
VERSION      BE448243.1   GI:9447820
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 399)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Tel: (301) 496-1550
             Email: Robert_Strausberg@nih.gov
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:1078440
             Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1..399
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3368836"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library is normalized. Library
                     was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT       99 a     101 c      98 g     101 t
ORIGIN

Query Match        74.5%;   Score 254.8;  DB 167;  Length 399;
Best Local Similarity  89.3%;   Pred. No. 4.3e-68;
```

```
Qy    1 gttccaggagattacagctccagccacaatgattgaggtggtttgcaacgcgtctagg 60
        ||||| ||| |||| |||| |||||||||||||||||||||||||||||||| || ||
Db   31 GTTCGGAGGATATTGGAGCTCCAGCCACAATGATTGAGGTGGTTTGCAACGACCGTCTCGG 90

Qy   61 aagaaagtccgcgttaagtgcaacacaccgatgacaccatcgggacttgaagaaactgat 120
        |||||||||||| ||||||||||||||||||||||||||||| ||||||||||||||| ||
Db   91 AAAGAAAGTCCGGTTAAGTGCAACACCGATGACACCATCGGCGACTTGAAGAAACTGAT 150

Qy  121 agcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatttt 180
        || || ||||||||||||| |||||| |||||||||||||||||||||||||||||||||
Db  151 AGCTGCTCAAACTGGCACCCGCTGGAACAAGATCGTTCTTAAAAAGTGTACACGATTTT 210

Qy  181 taaggaccatgtatctctgggagattagaaatccacgatgggatgaacctggagctta 240
        |||||||||| || ||||||||||||| |||||||||||||||||||||||||||| ||
Db  211 TAAGGACCACCGTGTCTCTGGGAGATTATGAAATCCACGATGGGATGAACCTGGAGCTTTA 270

Qy  241 ttaccagtagagggggaattcctccacct------tgcccaacctgcttcctctctccc 292
        ||||||||||||||| |||||| |||||        ||||| ||||| |||||||||||
Db  271 TTACCAGTAGAGGGGGGATTCCTTCCTCCGCCCTGCTCTGCCCCTGCCCCTCTCCC 330

Qy  293 atggctcatttaacactgttgtagatgctcattttt 328
        || |||||| |||||||||||||||||||||||||||
Db  331 AT-CCTCATCTGCACACTGGTGTGTAGATGCTCATTTTT 365

RESULT 10
BE197360/c
LOCUS        BE197360        398 bp    mRNA                 EST      26-JUN-2000
DEFINITION   ug69c04.x1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
             IMAGE:1547622 3' similar to WP:F46F11.4 CE10602 ;, mRNA sequence.
ACCESSION    BE197360
VERSION      BE197360.1   GI:8709529
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 398)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Other ESTs: ug69c04.y1
             Contact: Robert Strausberg, Ph.D.
             Tel: (301) 496-1550
             Email: Robert_Strausberg@nih.gov
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:952970.
FEATURES             Location/Qualifiers
     source          1..398
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1547622"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library is normalized. Library
                     was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      102 a     96 c      99 g     101 t
ORIGIN

Query Match        74.5%;   Score 254.8;  DB 164;  Length 398;
Best Local Similarity  89.3%;   Pred. No. 4.3e-68;
Matches 300;  Conservative 0;  Mismatches 27;  Indels 9;  Gaps 2;
```

Matches 300: Conservative 0; Mismatches 27; Indels 9; Gaps 2;

```
Qy    1 gttccaggagattacagctccagccacaatgattgaggtggtttgcaacgaccgtctagg  60
        ||| ||| ||  ||||||||||||||||||||||||||||||||||||||||| ||
Db   43 GTTCGGAGGATATTGGAGCTCCAGCCACAATGATTGAGGTGGTTTGCAACGACCGTCTCGG 102

Qy   61 aaagaaagtccgcgttaagtgcaacacaccgatgacacaccgggagacttgaagaaactgat 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db  103 AAAGAAAGTCCGCGTTAAGTGCAACACCGATGACACCGACTTGAAGAAACTGAT 162

Qy  121 acgcgcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgagtttt 180
        || |||| |||||||| |||||||||||||||||||||||||||||||||||||||||
Db  163 AGCTGCTCAAACTGGCACCCGCTGGAACAAGATCGTTCTTAAAAAGTGGTACACGAGTTTT 222

Qy  181 taaggaccatgtatctctggagattatgaaatccacgatggatgaacctggagcttta 240
        |||||||||| || |||||||||||||||||||||||||||||||||||||||||||
Db  223 TAAGGACCACGTGTCTCTGGAGATTATGAAATCCACGATGGATGAACCTGGAGCTTTA 282

Qy  241 ttaccagtagaggggaattcctccacct-------tgccaaccttgctttcctcctccc 292
        |||||||||||||||||||||||||||        || ||||||||||||||||||||
Db  283 TTACCAGTAGAGGGGGATTCCTTCCTCTCCGCCCTGCTCCTCCGCCCCTGCTCCTCTCCC 342

Qy  293 atggctcatttaacactgttgtagagtgctcatttt 328
        || ||||| ||||||||||||||||||||||||
Db  343 AT-CCTCATCTGACACTGGTGTAGATGTCATTTTT 377
```

RESULT 12

AA050441

LOCUS   AA050441   403 bp   mRNA   EST   09-SEP-1996
DEFINITION mj12d12.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:475895 5', mRNA sequence.
ACCESSION   AA050441
VERSION   AA050441.1   GI:1530112
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 403)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE   The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
  COMMENT   Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:286639
            Putative full length read
            vector to vector length is 405
            Seq primer: -28M13 rev2 from Amersham.
FEATURES   Location/Qualifiers
     source   1..403
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:475895"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGACGGCGCCGCCGGAAATTTTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo. "

BASE COUNT   98 a   100 c   98 g   107 t
ORIGIN

Query Match   74.5%;   Score 254.8; DB 1; Length 403;
Best Local Similarity 89.3%;   Pred. No. 4.3e-68;
Matches 300; Conservative 0; Mismatches 27; Indels 9; Gaps   2;

```
Qy    1 gttccaggagattacagctccagccacaatgattgaggtggtttgcaacgaccgtctagg  60
        ||| ||| ||  ||||||||||||||||||||||||||||||||||||||||| ||
Db   24 GTTCGGAGGATATTGGAGCTCCAGCCACAATGATTGAGGTGGTTTGCAACGACCGTCTCGG  83

Qy   61 aaagaaagtccgcgttaagtgcaacacaccgatgacacaccgggagacttgaagaaactgat 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   84 AAAGAAAGTCCGCGTTAAGTGCAACACCGATGACACCGACTTGAAGAAACTGAT 143

Qy  121 agcgcgcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgagtttt 180
        ||| |||| |||||||| |||||||||||||||||||||||||||||||||||||||||
Db  144 AGCTGCTCAAACTGGCACCCGCTGGAACAAGATCGTTCTTAAAAAGTGGTACACGAGTTTT 203

Qy  181 taaggaccatgtatctctggagattatgaaatccacgatggatgaacctggagcttta 240
        |||||||||| || |||||||||||||||||||||||||||||||||||||||||||
Db  204 TAAGGACCACGTGTCTCTGGAGATTATGAAATCCACGATGGATGAACCTGGAGCTTTA 263

Qy  241 ttaccagtagaggggaattcctccacct-------tgccaaccttgctttcctcctccc 292
        |||||||||||||||||||||||||||        || ||||||||||||||||||||
Db  264 TTACCAGTAGAGGGGGATTCCTTCCTCTCCGCCCTGCTCCTCCGCCCCTGCTCCTCTCCC 323

Qy  293 atggctcatttaacactgttgtagagtgctcatttt 328
        || ||||| ||||||||||||||||||||||||
Db  324 AT-CCTCATCTGACACTGGTGTAGATGTCATTTTT 358
```

RESULT 13

AA462867

LOCUS   AA462867   406 bp   mRNA   EST   10-JUN-1997
DEFINITION vf9ih07.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:851197 5', mRNA sequence.
ACCESSION   AA462867
VERSION   AA462867.1   GI:2187758
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
            1 (bases 1 to 406)
REFERENCE
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE   The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
  COMMENT   Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:503349

```
          Seq primer: -28m13 rev2 ET from Amersham
          High quality sequence stop: 379.
FEATURES          Location/Qualifiers
     source       1..406
                  /organism="Mus musculus"
                  /strain="C57BL/6J"
                  /db_xref="taxon:10090"
                  /clone="IMAGE:851197"
                  /clone_lib="Soares_mammary_gland_NbMMG"
                  /sex="male"
                  /tissue_type="mammary gland"
                  /dev_stage="4 weeks"
                  /lab_host="DH10B"
                  /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                  ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                  RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                  primer [5'
                  TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGTTTTTTTTTTTTTTTTTTTT
                  T 3']; double-stranded cDNA was ligated to Eco RI
                  adaptors (Pharmacia), digested with Not I and cloned into
                  the Not I and Eco RI sites of the modified pT7T3 vector.
                  RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                  constructed and normalized by Bento Soares and M.Fatima
                  Bonaldo."
BASE COUNT     124 a     94 c     90 g     98 t
ORIGIN

Query Match           74.5%;     Score 254.8; DB 7;   Length 406;
Best Local Similarity 89.3%;     Pred. No. 4.3e-68;
Matches 300; Conservative    0; Mismatches 27;    Indels   9;   Gaps   2;

Qy    1 gttccaggagattacagctccagccacaatgattgaggtggtttgcaacgaccgtctagg 60
          |||| |||| || ||||||||||||||||||||||||||||||||||||||||||| ||
Db    9 GTTCGAGGATATTGGAGCTCCAGCCACAATGATTGAGGTGGTTTGCAACGACCGTCTCGG 68

Qy   61 aaagaagtccgcgttaagtgcaacaccgatgcacaccgtcggggacttgaagaaactgat 120
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69 AAAGAAAGTCCGCGTTAAGTGCAACACCGATGCACACCGTCGGGGACTTGAAGAAACTGAT 128

Qy  121 agcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatttt 180
          || |||| |||||||||| ||| |||||| ||||||||||||||||||||||||||||||
Db  129 AGCTGCTCAAACTGGCACCCGCTGGAACAAGATCGTTCTTAAAAAGTGGTACACGATTTT 188

Qy  181 taaggaccatgtatctctgggagattatgaaatccacgatgggatgaacctggagcttta 240
          ||||||||| || |||||||||||||||||||| ||||||||||||||||||||||||||
Db  189 TAAGGACCACCGTGTCTCTGGGAGATTATGAAATCCACGATGGGATGAACCTGGAGCTTTA 248

Qy  241 ttaccagtagaggggaattcctccacct--------tgcccaaccttgcttcctctccc 292
          ||||||||||||||||||||||||||||        | |||||||||||||||||||||
Db  249 TTACCAGTAGAGGGGAATTCCTCCTCCCTCGCCCCTCGCCCCTGCCCCTCGCCCCTCTCCC 308

Qy  293 atggctcatttaacactgttgtagatgctcatttt 328
          || ||||| |||||||| ||||| |||||||||||
Db  309 AT-CCTCATCTGACACTGGTTGTAGATGGTGCATTTT 343

RESULT 14
LOCUS       W83387          410 bp    mRNA              EST       12-SEP-1996
DEFINITION  mf28c03.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:406372 5', mRNA sequence.
ACCESSION   W83387
VERSION     W83387.1  GI:1540754
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 410)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
```

```
TITLE      The WashU-HHMI Mouse EST Project
JOURNAL    Unpublished (1996)
COMMENT    On Sep 12, 1996 this sequence version replaced gi:1394358.
           Contact: Marra M/Mouse EST Project
           WashU-HHMI Mouse EST Project
           Washington University School of MedicineP
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: mouseest@watson.wustl.edu
           This clone is available royalty-free through LLNL ; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           MGI:250140
           Seq primer: -28M13 rev2 from Amersham
           High quality sequence stop: 378.
FEATURES          Location/Qualifiers
     source       1..410
                  /organism="Mus musculus"
                  /strain="C57BL/6J"
                  /db_xref="taxon:10090"
                  /clone="IMAGE:406372"
                  /clone_lib="Soares mouse embryo NbME13.5 14.5"
                  /sex="unknown"
                  /tissue_type="embryo"
                  /dev_stage="13.5-14.5dpc total fetus"
                  /lab_host="DH10B"
                  /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                  polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                  was primed with a Not I - oligo(dT) primer [5'
                  TGTTACCAATCTGAAGTGGGAGCGGCCGCGAAATTTTTTTTTTTTTTTTTTTT
                  T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                  14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
                  State Univ., from 2 ]; double-stranded cDNA was ligated to
                  Eco RI adaptors (Pharmacia), digested with Not I and
                  cloned into the Not I and Eco RI sites of the modified
                  pT7T3 vector. Library went through one round of
                  normalization, and was constructed by Bento Soares and
                  M.Fatima Bonaldo. "
BASE COUNT     96 a     105 c     103 g     106 t
ORIGIN

Query Match           74.5%;     Score 254.8; DB 158;  Length 410;
Best Local Similarity 89.3%;     Pred. No. 4.3e-68;
Matches 300; Conservative    0; Mismatches 27;    Indels   9;   Gaps   2;

Qy    1 gttccaggagattacagctccagccacaatgattgaggtggtttgcaacgaccgtctagg 60
          |||| |||| || ||||||||||||||||||||||||||||||||||||||||||| ||
Db   62 GTTCGAGGATATTGGAGCTCCAGCCACAATGATTGAGGTGGTTTGCAACGACCGTCTCGG 121

Qy   61 aaagaagtccgcgttaagtgcaacaccgatgcacaccgtcggggacttgaagaaactgat 120
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  122 AAAGAAAGTCCGCGTTAAGTGCAACACCGATGCACACCGTCGGGGACTTGAAGAAACTGAT 181

Qy  121 agcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatttt 180
          || |||| |||||||||| ||| |||||| ||||||||||||||||||||||||||||||
Db  182 AGCTGCTCAAACTGGCACCCGCTGGAACAAGATCGTTCTTAAAAAGTGGTACACGATTTT 241

Qy  181 taaggaccatgtatctctgggagattatgaaatccacgatgggatgaacctggagcttta 240
          ||||||||| || |||||||||||||||||||| ||||||||||||||||||||||||||
Db  242 TAAGGACCACCGTGTCTCTGGGAGATTATGAAATCCACGATGGGATGAACCTGGAGCTTTA 301

Qy  241 ttaccagtagaggggaattcctccacct--------tgcccaaccttgcttcctctccc 292
          ||||||||||||||||||||||||||||        | |||||||||||||||||||||
Db  302 TTACCAGTAGAGGGGAATTCCTCCTCCCTCGCCCCTCGCCCCTGCCCCTCGCCCCTCTCCC 361

Qy  293 atggctcatttaacactgttgtagatgctcatttt 328
          || ||||| |||||||| ||||| |||||||||||
Db  362 AT-CCTCATCTGACACTGGTTGTAGATGGTGCATTTT 396
```

```
Qy    241  ttaccagtagaggggaattcctccacct------tgccaacttgctttcctctccc  292
           ||||||||||||||||||  |  ||  |||   |||||||||||||
Db    295  TTACCAGTAGAGGGGAATTCCTTCCTCCTCCTCCGCCCTGCTCTGCCCTGCCCCTCCTCTCCC  354

Qy    293  atggctcattaacactgttgtagatgctcatttt  328
           ||  |||||  | ||||||| |||||||| ||||||||
Db    355  AT-CCTCATCTGACACTGGTGTAGATGGTCATTTTT  389

Search completed: June 13, 2001, 19:58:50
Job time: 5023 sec
```

RESULT 15
LOCUS   W11593   411 bp   mRNA   EST   02-OCT-1997
DEFINITION ma90a06.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:317938 5', mRNA sequence.
ACCESSION   W11593
VERSION   W11593.1   GI:1285895
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 411)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL.; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI: 208554
FEATURES   Location/Qualifiers
     source   1..411
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:317938"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
BASE COUNT   101 a   105 c   101 g   104 t
ORIGIN

```
Query Match            74.5%;  Score 254.8;  DB 158;  Length 411;
Best Local Similarity  89.3%;  Pred. No. 4.3e-68;
  Matches 300;  Conservative   0;  Mismatches  27;  Indels   9;  Gaps   2;

Qy      1  gttccaggagattacagctccagccaccaatgattgaggtggttgcaacgaccgtctagg   60
           ||||  ||||||||||||||||||||||||||||||||||||||  |||||||||||||| ||
Db     55  GTTCGAGGAGATATTGGAGCTCCAGCCACCAATGATTGAGGTGGTTGCAACGACCGTCTCGG  114

Qy     61  aaagaaagtccgcgttaagtgcaacaccgatgacacccatcgggacttgaagaaactgat  120
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    115  AAAGAAAGTCCGCGTTAAGTGCAACACCGATGACACCATCGGCGACTTGAAGAAAACTGAT  174

Qy    121  agcggcccaaactggcactcgttggaataagatcgttcttaaaagtggtaccgattttt  180
           ||||| |||||||||||  ||||| ||  ||||||||||||||||||||||||||||||||
Db    175  AGCTGCTCAAACTGGCACCCCGCTGGAACAAGATCGTTCTTAAAAGTGGTACCACGATTTT  234

Qy    181  taaggaccatgtatctcctggagattatgaaatccacgatggatgaacctggagcttta  240
           ||||||||||||||  |||||||||||||||||  ||||||||||||||||||||||||||
Db    235  TAAGGACCACCGTGTCTCTGGGAGATTATGAAATCCACGATGGGATGAACCTGGAGCTTTA  294
```

```
               GenCore version 4.5
          Copyright (c) 1993 - 2000  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:   June 13, 2001, 18:39:17 ; Search time 86.89 Seconds
                                     (without alignments)
                                     687.238 Million cell updates/sec

Title:          US-09-331-930A-1
Perfect score:  342
Sequence:       1 gttccaggagattacagctc.........attttttgttaagtgtact 342

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

arched:   302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters:   605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Issued_Patents_NA:*
    1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
    2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
    3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
    4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
    5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
    6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	35.4	10.4	12492	6	5206163-2	Patent No. 5206163
2	32.2	9.4	3840	3	US-08-462-092-1	Sequence 1, Appli
3	32.2	9.4	3840	3	US-08-746-822-1	Sequence 1, Appli
4	32.2	9.4	3840	3	US-09-094-350-1	Sequence 1, Appli
5	32.2	9.4	3840	6	5510474-1	Patent No. 5510474
6 c	31.8	9.3	1757	1	US-08-313-075A-49	Sequence 49, Appl
7	30.6	8.9	4245	5	PCT-US96-05320A-516	Sequence 516, App
8 c	30	8.8	3571	4	US-09-280-799-132	Sequence 132, App
9 c	30	8.8	3571	4	US-09-277-020-9	Sequence 9, Appli
10	29.6	8.7	581	3	US-08-557-309B-22	Sequence 22, Appl
11	29.6	8.7	581	3	US-08-834-306-22	Sequence 22, Appl
12 c	29	8.5	7218	1	US-08-232-463-14	Sequence 14, Appl
13	28.4	8.3	707	2	US-08-726-306A-22	Sequence 22, Appl
14	28.4	8.2	7218	1	US-08-232-463-14	Sequence 3, Appli
15	28.2	8.2	1800	1	US-08-848-252-3	Sequence 3, Appli
16	28.2	8.2	2022	2	US-08-505-486-96	Sequence 96, Appl
17	28.2	8.2	2022	2	US-08-801-028-96	Sequence 96, Appl
18	28.2	8.2	2022	3	US-09-340-154-96	Sequence 96, Appl
19	28.2	8.2	2022	5	PCT-US95-09338-96	Sequence 96, Appl
20	28.2	8.2	2022	5	PCT-US95-09339-96	Sequence 96, Appl
21	28.2	8.2	2127	2	US-08-505-486-95	Sequence 95, Appl
22	28.2	8.2	2127	3	US-08-801-028-95	Sequence 95, Appl
23	28.2	8.2	2127	3	US-09-340-154-95	Sequence 95, Appl
24	28.2	8.2	2127	5	PCT-US95-09338-95	Sequence 95, Appl
25	28.2	8.2	2127	5	PCT-US95-09339-95	Sequence 95, Appl
26	28.2	8.2	2252	4	US-08-462-467B-13	Sequence 13, Appl
27	28	8.2	554	3	US-08-462-146-16	Sequence 16, Appl
28	28	8.2	554	3	US-09-360-220-16	Sequence 16, Appl
29	28	8.2	2241	3	US-08-840-146-1	Sequence 1, Appli
30	28	8.2	2241	2	US-09-360-220-1	Sequence 1, Appli
31	27.8	8.1	3621	2	US-09-019-201A-1	Sequence 1, Appli
32	27.6	8.1	366	4	US-08-881-771A-7	Sequence 7, Appli
33	27.6	8.1	1096	4	US-08-881-771A-5	Sequence 5, Appli
34	27.6	8.0	2105	4	US-09-088-425-3	Sequence 3, Appli
35	27.4	8.0	231	1	US-09-450-834-1	Sequence 1, Appli
36	27.4	8.0	831	2	US-08-450-834-5	Sequence 5, Appli
37 c	27.2	8.0	35100	2	US-08-770-379-17	Sequence 17, Appl
38 c	27.2	8.0	35100	4	US-08-757-669A-17	Sequence 17, Appl
39	27	7.9	1817	3	US-08-564-496C-14	Sequence 14, Appl
40 c	27	7.9	3349	2	US-08-920-234-1	Sequence 1, Appli
41 c	26.8	7.8	2664	2	US-08-942-819-1	Sequence 1, Appli
42 c	26.6	7.8	285	6	5487983-3	Patent No. 5487983
43 c	26.6	7.8	4481	4	US-08-867-611-1	Sequence 1, Appli
44 c	26.6	7.8	4481	5	PCT-US92-06965A-6	Sequence 6, Appli
45 c	26.6	7.8	5600	4	US-08-867-611-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
5206163-2
;Patent No. 5206163
;APPLICANT: RENARD, ANDRE;DINA, DINO;MARTIAL, JOSEPH
;TITLE OF INVENTION: DNA ENCODING BOVINE DIARRHEA
;VIRUS PROTEIN
;NUMBER OF SEQUENCES: 3
;CURRENT APPLICATION DATA:
;APPLICATION NUMBER: US/07/550,816
;FILING DATE: 06-JUL-1990
;PRIOR APPLICATION DATA:
;APPLICATION NUMBER: 331,037
;FILING DATE: 29-MAR-1989
;APPLICATION NUMBER: 752,981
;FILING DATE: 08-JUL-1985
;SEQ ID NO:2:
;LENGTH:12492
5206163-2

Query Match             10.4%;   Score 35.4;   DB 6;   Length 12492;
Best Local Similarity   46.8%;   Pred. No. 0.017;
Matches  111;  Conservative   0;  Mismatches 126;  Indels   0;  Gaps   0;

Qy      28  aatgatgaggtggttgcaacgaccgtctaggaaagaaagtccgcgttaagtgcaacac 87
            || || ||||| || |||||||||| ||||  ||| |||| ||| || |||| |||||
Db    5148  aaaaatgcagatcttcgtgaaaacctgaccggcaagacccatcacccgagggagagagc 5207

Qy      88  cgatgaccatcgggggacttgaagaaactgatagcggccaaactggcactcgttggaa 147
            | |||||||||||| || ||||| ||| || |||| ||  |  ||| || |||| ||
Db    5208  cagtgacaccatcgagaacgtgaaggccaagatccagataaggaaggcattccccctga 5267

Qy     148  taagatcgttctaaaaagtggtacacgatttttaaggaccatgtatctctgggagatta 207
            ||      || | ||||| ||| |   ||  ||| |   || ||| ||  ||||  ||||
Db    5268  ccagcagggctcatcttgccgtgcaagcagcggaagatggccgcctcttcttttctgatta 5327

Qy     208  tgaaatccacgatggatgaactcgaccctggagcttatttattaccagtagagggaattcctcc 264
            || ||| | || ||| ||| || | ||| ||| || |  ||| | || ||| |  |||| ||
Db    5328  caacatccagaaagagtcgacccctgacaccggtcctccgtctcgaggggtagtgggcc 5384

RESULT 2
US-08-462-092-1
; Sequence 1, Application US/08462092
; Patent No. 5614399
;GENERAL INFORMATION:
;APPLICANT: Peter H. Quail
;APPLICANT: Alan H. Christensen
;APPLICANT: Howard P. Hershey
;APPLICANT: Robert A. Sharrock
```

; APPLICANT: Thomas D. Sullivan
; TITLE OF INVENTION: PLANT UBIQUITIN PROMOTER SYSTEM
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jeff Lloyd
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,092
; FILING DATE: 5-JUNE-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; REGISTRATION NUMBER: 35,589
; REFERENCE/DOCKET NUMBER: 08/462,092
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3840 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1993..3591
US-08-462-092-1

Query Match　　　9.4%;　Score 32.2;　DB 1;　Length 3840;
Best Local Similarity 47.7%;　Pred. No. 0.13;
Matches　94;　Conservative　0;　Mismatches 103;　Indels 103;　Gaps　0;

Qy　59 ggaaagaaagtccgcgttaagtgcaacaccgatgacaccggggacttgaagaaactg 118
       ||·|||| ||·||·||·|||·||·|·|||||||||||·|·|||||||||·||·||||
Db 3160 GGCAAGACCATCACCTTGGGAGTGGAGAGCTTCGACACCATTGACAATGTGAAGGCCAAG 3219

Qy 119 atagcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtaccacgatt 178
       || ||·||||·|·|||||·|·||·||·||·|||||·|||·||||||·||·||||·|||·||
Db 3220 ATCCAGGACAAGGAGGGCAATCCCCCAGACCAGCGTCTGATCTTTGCGGGCAAGCAG 3279

Qy 179 tttaaggaccatgtatctctgggagattatgaaatccacgatgggatgaacctgggagctt 238
       ||·|| |||||·|·|·|·|·|||·||·||·||·||||·|·|·|·||·|||·|||·|||·|||
Db 3280 CTGGAGGATGGCCGCACTCTCGCGGACTACAACATCCAGAAGGAGGAGCACCCTTCACCTT 3339

Qy 239 tattaccagtagagggg 255
       ·|·|·|·|||·|||·||
Db 3340 GTTCTCCGCCTCAGAGG 3356

RESULT 3
US-08-746-822-1
; Sequence 1, Application US/08746822
; Patent No. 6020190
; GENERAL INFORMATION:
; APPLICANT: Peter H. Quail
; APPLICANT: Alan H. Christensen
; APPLICANT: Howard P. Hershey
; APPLICANT: Robert A. Sharrock
; APPLICANT: Thomas D. Sullivan
; TITLE OF INVENTION: PLANT UBIQUITIN PROMOTER SYSTEM
; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jeff Lloyd
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,822
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,092
; FILING DATE: 5-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; REGISTRATION NUMBER: 35,589
; REFERENCE/DOCKET NUMBER: 08/462,092
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3840 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1993..3591
US-08-746-822-1

Query Match　　　9.4%;　Score 32.2;　DB 3;　Length 3840;
Best Local Similarity 47.7%;　Pred. No. 0.13;
Matches　94;　Conservative　0;　Mismatches 103;　Indels 103;　Gaps　0;

Qy　59 ggaaagaaagtccgcgttaagtgcaacaccgatgacaccggggacttgaagaaactg 118
       ||·|||| ||·||·||·|||·||·|·|||||||||||·|·|||||||||·||·||||
Db 3160 GGCAAGACCATCACCTTGGGAGTGGAGAGCTTCGACACCATTGACAATGTGAAGGCCAAG 3219

Qy 119 atagcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtaccacgatt 178
       || ||·||||·|·|||||·|·||·||·||·|||||·|||·||||||·||·||||·|||·||
Db 3220 ATCCAGGACAAGGAGGGCAATCCCCCAGACCAGCGTCTGATCTTTGCGGGCAAGCAG 3279

Qy 179 tttaaggaccatgtatctctgggagattatgaaatccacgatgggatgaacctgggagctt 238
       ||·|| |||||·|·|·|·|·|||·||·||·||·||||·|·|·|·||·|||·|||·|||·|||
Db 3280 CTGGAGGATGGCCGCACTCTCGCGGACTACAACATCCAGAAGGAGGAGCACCCTTCACCTT 3339

Qy 239 tattaccagtagagggg 255
       ·|·|·|·|||·|||·||
Db 3340 GTTCTCCGCCTCAGAGG 3356

RESULT 4
US-09-094-350-1
; Sequence 1, Application US/09094350
; Patent No. 6054574
; GENERAL INFORMATION:
; APPLICANT: Peter H. Quail
; APPLICANT: Alan H. Christensen
; APPLICANT: Howard P. Hershey
; APPLICANT: Robert A. Sharrock
; APPLICANT: Thomas D. Sullivan
; TITLE OF INVENTION: PLANT UBIQUITIN PROMOTER SYSTEM
; NUMBER OF SEQUENCES: 2

```
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Jeff Lloyd
;     STREET:  2421 N.W. 41st Street, Suite A-1
;     CITY:  Gainesville
;     STATE:  FL
;     COUNTRY:  USA
;     ZIP:  32606
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/09/094,350
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/746,822
;     FILING DATE:
;     APPLICATION NUMBER:  US/08/462,092
;     FILING DATE:  5-JUNE-1995
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Lloyd, Jeff
;     REGISTRATION NUMBER:  35,589
;     REFERENCE/DOCKET NUMBER:  08/462,092
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  904-375-8100
;     TELEFAX:  904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  3840 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     MOLECULE TYPE:  DNA (genomic)
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  1993..3591
US-09-094-350-1

Query Match               9.4%;  Score 32.2; DB 3; Length 3840;
Best Local Similarity   47.7%;  Pred. No. 0.13;
Matches   94; Conservative    0; Mismatches 103; Indels    0; Gaps    0;

     Qy    59 ggaaagaaagtccgcgttaagtgcaacaccgatgacaccatcgggagactgaagaaactg 118
              || |||| ||-|-|| ||| | || |||||||||||||||| | |||| -| ||||
     Db  3160 GGCAAGACCATCACCTTGGAGGTGGAGCTCTGACACCATTGACACCATTGACAATGTGAAGGCCAAG 3219

     Qy   119 atagcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatt 178
              || | ||||| || | ||||| ||-| | |-|||| |-|||| -| ||||
     Db  3220 ATCCAGGACAAGGAGGGCATTCCCCCAGACCAGGCGTCTGATCTTTGCGGCAAGCAG 3279

     Qy   179 tttaaggaccatgtatctctgggagattatgaaatccacgatgggatgaacctggagctt 238
              -| |||| ||||| |-|| || | |||| ||-||| -| ||||| -| |||
     Db  3280 CTGGAGGATGGCCGCACTCTCGCGGACTACAACATCCAGAAGGAGGAGCACCCCTTCACCTT 3339

     Qy   239 tattaccagtagagggg 255
              -| |||| -| ||||
     Db  3340 GTTCTCCGCCTCAGAGG 3356

RESULT 5
5510474-1
;Patent No. 5510474
; APPLICANT:  QUAIL, PETER H.;CHRISTENSEN, ALAN H.;HERSHEY,
;HOWARD P.;SHARROCK, ROBERT A.;SULLIVAN, THOMAS D.
; TITLE OF INVENTION:  PLANT UBIQUITIN PROMOTER SYSTEM
; NUMBER OF SEQUENCES: 12
; CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/296,268
;     FILING DATE:  25-AUG-1994
```

```
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  191,134
;     FILING DATE:  03-FEB-1994
;     APPLICATION NUMBER:  76,363
;     FILING DATE:  11-JUN-1993
;     APPLICATION NUMBER:  670,496
;     FILING DATE:  15-MAR-1991
;     APPLICATION NUMBER:  194,824
;     FILING DATE:  17-MAY-1988
; SEQ ID NO:1:
;   LENGTH: 3840
5510474-1

Query Match               9.4%;  Score 32.2; DB 6; Length 3840;
Best Local Similarity   47.7%;  Pred. No. 0.13;
Matches   94; Conservative    0; Mismatches 103; Indels    0; Gaps    0;

     Qy    59 ggaaagaaagtccgcgttaagtgcaacaccgatgacaccatcgggagactgaagaaactg 118
              || |||| ||-|-|| ||| | || |||||||||||||||| | |||| -| ||||
     Db  3160 ggcaagaccatcaccttggaggtggagctctgacaccattgacaccattgacaatgtgaaggccaag 3219

     Qy   119 atagcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatt 178
              || | ||||| || | ||||| ||-| | |-|||| |-|||| -| ||||
     Db  3220 atccaggacaaggagggcattcccccagaccagcgtctgatctttgcgggcaagcag 3279

     Qy   179 tttaaggaccatgtatctctgggagattatgaaatccacgatgggatgaacctggagctt 238
              -| |||| ||||| |-|| || | |||| ||-||| -| ||||| -| |||
     Db  3280 ctggaggatggccgcactctcgcggactacaacatccagaaggaggagcacccctcacctt 3339

     Qy   239 tattaccagtagagggg 255
              -| |||| -| ||||
     Db  3340 gttctccgcctcagagg 3356

RESULT  6
US-08-313-075A-49/c
; Sequence 49, Application US/08313075A
; Patent No. 5639870
; GENERAL INFORMATION:
;   APPLICANT:  Holton, Timothy A.
;   APPLICANT:  Cornish, Edwina C.
;   APPLICANT:  Tanaka, Yoshikazu
;   TITLE OF INVENTION:  GENETIC SEQUENCES ENCODING FLAVONOID
;   TITLE OF INVENTION:  PATHWAY ENZYMES AND USES THEREFOR
;   NUMBER OF SEQUENCES:  58
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Scully, Scott, Murphy & Presser
;     STREET:  400 Garden City Plaza
;     CITY:  Garden City
;     STATE:  New York
;     COUNTRY:  U.S.A.
;     ZIP:  11530
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/313,075A
;     FILING DATE:  30-NOV-1994
;     CLASSIFICATION:  800
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  AU PL 1538/92
;     FILING DATE:  27-MAR-1992
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  AU PL 6698/93
;     FILING DATE:  07-JAN-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  AU PCT/AU93/00127
;     FILING DATE:  25-MAR-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  DiGiglio, Frank S.
```

;   REGISTRATION NUMBER: 31,346
;   REFERENCE/DOCKET NUMBER: 9433
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (516) 742-4343
;   TELEFAX: (516) 742-4366
;   TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 49:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1757 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 35..1522
US-08-313-075A-49

Query Match   9.3%; Score 31.8; DB 1; Length 1757;
Best Local Similarity 48.2%; Pred. No. 0.11;
Matches 119; Conservative   0; Mismatches 127; Indels   1; Gaps   1;

Qy   7 ggagattacagctccagccacaatgattgaggtggttgcaacgaccgtccaggaaagaa 66
            || || ||||| ||| || ||||| || ||||| | |||| |
Db   685 GGGTTTTCCAGCCGTCAACCATAATGTTCCAAACCAATTCCTTAAACTCTTTAGCAGAATC 626

Qy   67 agtccgcgttaagtgcaaccaccgatgacaccatcggggacttgaagaaactgatagcggc 126
            || ||| |||| ||| ||| | || || || || |
Db   625 AGAAAACGGATTAGTCAAATCTTTAGAGAAAATGGTGTTCGGATAGCAAATTCAAAGTAGT 566

Qy   127 ccaaactgcactcgttgga-ataagatcgttcttaaaaagtggtacacgatttttaagg 185
            || | ||| ||| | |||| | ||| | || ||
Db   565 TCCAAAGTTGCTCTTCCTATATCCACTCGTTGGCCCATATCTTGGCCACACTTTTGACAATA 506

Qy   186 accatgtatctctggagattatgaaatccacgatgggatggaacctttattacc 245
            || ||| || |||| ||| | ||||| |
Db   505 ATCAATTAACTCTTGGACCTTTTAGACCTCAGATGTTGATTACCATCAAGCTTGTTACC 446

Qy   246 agtagag 252
            || ||
Db   445 AGAAAAG 439

RESULT 7
PCT-US96-05320A-516
; Sequence 516, Application PC/TUS9605320A
; GENERAL INFORMATION:
;   APPLICANT: Human Genome Sciences
;   APPLICANT: 9410 Key West Avenue
;   APPLICANT: Rockville, MD 20850
;   APPLICANT: United States of America
;   APPLICANT: Johns Hopkins University
;   APPLICANT: 720 Rutland Avenue
;   APPLICANT: Baltimore, MD 21205
;   APPLICANT: United States of America
;   APPLICANT: Mark D. Adams
;   APPLICANT: Owen White
;   APPLICANT: Hamilton O. Smith
;   APPLICANT: J. Craig Venter
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Sterne, Kessler, Goldstein & Fox
;   STREET: 1100 New York Avenue, Suite 600
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20003-3934
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;   COMPUTER: HP Vectra 486/33
;   OPERATING SYSTEM: MSDOS version 6.2

;   SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US96/05320A
;   FILING DATE: April22, 1996
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/476,102
;   FILING DATE: June 7, 1995
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/487,429
;   FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
;   NAME: Eric K. Steffe
;   REGISTRATION NUMBER: 36,688
;   REFERENCE/DOCKET NUMBER: 1488.014PC01
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 371-2600
;   TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 516:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 4245 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
PCT-US96-05320A-516

Query Match   8.9%; Score 30.6; DB 5; Length 4245;
Best Local Similarity 55.0%; Pred. No. 0.52;
Matches 60; Conservative   0; Mismatches 49; Indels   0; Gaps   0;

Qy   103 ggacttgaagaaactgatagcggccaaactggcactcgttggaataagatcgttcttaa 162
            ||| ||| ||| ||||| |||| |||| |||| |||||
Db   1054 GGTCGTTCGTAATCACTGTTGGCCCATATCTTGACTTACCACCATGTGTTTACCGAAG 1113

Qy   163 aaagtggtacacgattttaaggaccacatgtatctctgggagattatgaa 211
            ||| ||| |||| | |||| || |||| | |||
Db   1114 AAAATGGCATTGGAATTATTCCGTCCGTTTATCTATGCGAAATTAGAAA 1162

RESULT 8
US-09-280-799-132/c
; Sequence 132, Application US/09280799
; Patent No. 6136603
; GENERAL INFORMATION:
;   APPLICANT: Dean, Nicholas M.
;   APPLICANT: Karras, James G
;   APPLICANT: McKay, Robert
; TITLE OF INVENTION: ANTISENSE MODULATION OF INTERLEUKIN-5 SIGNAL
; TITLE OF INVENTION: TRANSDUCTION
; FILE REFERENCE: ISPH-0340
; CURRENT APPLICATION NUMBER: US/09/280,799
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 132
; LENGTH: 3571
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Synthetic
US-09-280-799-132

Query Match   8.8%; Score 30; DB 4; Length 3571;
Best Local Similarity 49.4%; Pred. No. 0.78;
Matches 78; Conservative   0; Mismatches 80; Indels   0; Gaps   0;

Qy   176 attttaaggaccatgtatctctggagattatgaaatccacgatggggatgaacctggag 235
            ||||| |||||| |||| | || |||| | || | || ||| |||| ||||| ||
Db   2600 AATTTTAAGACCATTTCTCAGTGAACCTGAACTAGAGATAACAATAATTGGCTAAACTGGTTG 2541

Qy   236 cttattaccagtagaggggaattcctccacctgcccaaccttgcttcctctccccatg 295

```
Db  2540 AACTACAACAATTCCCATGATCCTCCTTGCTTTATCTCCCAACTCTGGGGTTATAGGAATGT 2481
             ||  ||    || ||                        | |||||| ||    |
Qy   296 gctcatttaacactgttgtagatgctcattttttgtt 333
             || ||   || |  |||| ||      |  | |||
Db  2480 GCTTCTTGCACTGTTTTGTTGTTGCTGCTGCTGTTGTT 2443

RESULT 9
US-09-277-020-9/c
; Sequence 9, Application US/09277020
; Patent No. 6210892
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  TITLE OF INVENTION: Alteration of Cellular Behavior by Antisense Modulation
;  TITLE OF INVENTION: of mRNA Processing
;  FILE REFERENCE: ISPH-0339
; CURRENT APPLICATION NUMBER: US/09/277,020
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 09/167,921
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;  LENGTH: 3571
;  TYPE: DNA
;  ORGANISM: Mus musculus
US-09-277-020-9

Query Match             8.8%;  Score 30;  DB 4;  Length 3571;
Best Local Similarity  49.4%;  Pred. No. 0.78;
Matches  78;  Conservative   0;  Mismatches  80;  Indels   0;  Gaps   0;

Qy   176 attttaaggaccatgtatctctgggagattatgaaatccacgatgggatgaacctggag 235
             | |||||||||||  |||||| |       || |||  | | ||||
Db  2600 AATTTTTAAGACCATTTCTCTCAGTGAACCTAGAGATAACATAATTGGCTAAACTGGTTG 2541
             | || ||        |  |        |    || ||  ||||
Qy   236 cttattaccagtagaggggaattcctccaacctgcccaacctgctttcctcccatg 295
             ||  |  ||  |          ||||     |||||   | |||||||
Db  2540 AACTACAACAATTCCCATGATCCTCCTTGCTTTATCTCCCAACTCTGGGGTTATAGGAATGT 2481
             ||  ||    || ||                        | |||||| ||    |
Qy   296 gctcatttaacactgttgtagatgctcattttttgtt 333
             || ||   || |  |||| ||      |  | |||
Db  2480 GCTTCTTGCACTGTTTTGTTGTTGCTGCTGCTGTTGTT 2443

RESULT 10
US-08-557-309B-22
; Sequence 22, Application US/08557309B
; Patent No. 591672
; GENERAL INFORMATION:
;  APPLICANT: Reed,Steven G.
;  APPLICANT: Skeiky, Yasir A.W.
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: Houghton, Raymond L.
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND    PREVENTION OF T
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
;  ADDRESSEE: SEED and BERRY LLP
;  STREET: 6300 Columbia Center, 701 Fifth Avenue
;  CITY: Seattle
;  STATE: Washington
;  COUNTRY: USA
;  ZIP: 98104-7092
; COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/557,309B
;  FILING DATE: 14-NOV-1995
```

```
;  CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;  NAME: Maki, David J.
;  REGISTRATION NUMBER: 31,392
;  REFERENCE/DOCKET NUMBER: 210121.422
; TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (206) 622-4900
;  TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 22:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 581 base pairs
;  TYPE: nucleic acid
;  STRANDEDNESS: single
;  TOPOLOGY: linear
US-08-557-309B-22

Query Match             8.7%;  Score 29.6;  DB 2;  Length 581;
Best Local Similarity  47.8%;  Pred. No. 0.37;
Matches  86;  Conservative   0;  Mismatches  94;  Indels   0;  Gaps   0;

Qy    59 ggaaagaaagtccggttaagtgcaacaccgatgacaccatcggggacttgaagaaactg 118
             || |||| || ||  | ||      ||||||||||   | ||||||| ||||| |||||
Db    72 GGCAAGAACGATCGCGCTCGAGGTGGAATCCAGCGACACCATTGAGAACGTGAAGGCGAAG 131
             |||| | || ||         | ||| || |  ||   ||   |    ||||
Qy   119 atagcggcccaaactgcactcgttgaataagatcgttcttaaaaagtggtacacgatt 178
             | |||   ||    ||  || | | |  ||   |   ||   ||   |||
Db   132 ATCCAGGACAAGGAGGGCATTCCGCCGGACCAGCAGCCGCCTGATCTTCGCTGGCAAGCAG 191
             |  ||||| ||||||||| ||||||
Qy   179 tttaaggaccatgtatctctgggagattatgaaatccacgatgggatgaacctggagctt 238
             | |||||      |||  |  ||||| |  || |||||||  || || |||  ||||||||
Db   192 CTGGAGGACGGCGCCGCACGCTCGCAGACTACACATCCAGAAGGAGTCCACGCTGCACCTT 251

RESULT 11
US-08-834-306-22
; Sequence 22, Application US/08834306
; Patent No. 6054135
; GENERAL INFORMATION:
;  APPLICANT: Reed,Steven G.
;  APPLICANT: Skeiky, Yasir A.W.
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: Houghton, Raymond L.
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND    PREVENTION OF
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
;  ADDRESSEE: SEED and BERRY LLP
;  STREET: 6300 Columbia Center, 701 Fifth Avenue
;  CITY: Seattle
;  STATE: Washington
;  COUNTRY: USA
;  ZIP: 98104-7092
; COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/834,306
;  FILING DATE: 15-APR-1997
;  CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
;  NAME: Maki, David J.
;  REGISTRATION NUMBER: 31,392
;  REFERENCE/DOCKET NUMBER: 210121.422C1
; TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (206) 622-4900
;  TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 22:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 581 base pairs
;  TYPE: nucleic acid
;  STRANDEDNESS: single
```

;   TOPOLOGY: linear
US-08-834-306-22

Query Match   8.7%; Score 29.6; DB 3; Length 581;
Best Local Similarity 47.8%; Pred. No. 0.37;
Matches 86; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```
Qy    59 ggaagaaagtccgcgttaagtgcaacaccgatgacaccatcgggagacttgaagaaactg 118
            || |||| || || | || |||||||| || |||||||||| ||||
Db    72 GGCAAGACGACTCGGCTCGAGGTGAATCCAGCGACACCATTGAGAACGTGAAGGCGAAG 131
Qy   119 atagcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatt 178
             |  |||||||| || ||  || |||| ||  ||    |
Db   132 ATCCAGGACAAGGAGGGCATTCCCGCGGACAGCAGCGGCCTGATCTTCGCTGGCAAGCAG 191
Qy   179 tttaaggaccatgtatctctgggagattatgaaatccacgatgggtgatgaacctgagctt 238
             | |||  | | || | | ||| | |  |  ||| | || |||
Db   192 CTGGAGGACGGCCGCCCACGCTCGCAGACTACAACATCCAGAAGGAGTCCACGCTGCACCTT 251
```

RESULT 12
08-232-463-14/c
;equence 14, Application US/08232463
Patent No. 5670367
; GENERAL INFORMATION:
;   APPLICANT: DORNER, F.
;   APPLICANT: SCHEIFLINGER, F.
;   APPLICANT: FALKNER, F. G.
;   TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;   NUMBER OF SEQUENCES: 52
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Foley & Lardner
;   STREET: 1800 Diagonal Road, Suite 500
;   CITY: Alexandria
;   STATE: VA
;   COUNTRY: USA
;   ZIP: 22313-0299
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/232,463
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/07/935,313
;   FILING DATE:
;   APPLICATION NUMBER: EP 91 114 300.6
;   FILING DATE: 26-AUG-1991
;   ATTORNEY/AGENT INFORMATION:
;   NAME: BENT, Stephen A.
;   REGISTRATION NUMBER: 29,768
;   REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703)836-9300
;   TELEFAX: (703)683-4109
;   TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 7218 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;   CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match   8.5%; Score 29; DB 1; Length 7218;
Best Local Similarity 4.0%; Pred. No. 2.8;

Matches 11; Conservative 146; Mismatches 116; Indels 0; Gaps 0;

```
Qy     6 aggagattacagctccagccacaatgattgaggtggttgcaacgaccgtctaggaaaga 65
                                                              ::::::::::
Db  1319 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1260
Qy    66 aagtccgcgttaagtgcaacaccgatgacaccatcgggagacttgaagaaactgatagcgg 125
                                                              ::: :::::
Db  1259 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1200
Qy   126 cccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgattttaagg 185
                                                              ::: :::
Db  1199 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1140
Qy   186 accatgtatctctgggagattatgaaatccacgatgggtgaacctgagctttattacc 245
                                                              ::: :::
Db  1139 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1080
Qy   246 agtagagggaattcctccaccttgccaacct 278
                        |  ||| |||| ||||
Db  1079 RRRRRRRRRRRRATCGCAAGCTCCCTCGACCT 1047
```

RESULT 13
US-08-726-306A-22
; Sequence 22, Application US/08726306A
; Patent No. 5958684
; GENERAL INFORMATION:
; APPLICANT: van Leeuwen, Frederik Willem
; APPLICANT: Burbach, Johannes Peter Henri
; APPLICANT: Grosveld, Franklin G.
; TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS
; NUMBER OF SEQUENCES: 189
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Banner & Witcoff, Ltd.
;   STREET: 1 Financial Center
;   CITY: Boston
;   STATE: MA
;   COUNTRY: US
;   ZIP: 02111
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/726,306A
;   FILING DATE: 02-Oct-1996
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: GB 95/20080.4
;   FILING DATE: 02-Oct-1995
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 60/009,832
;   FILING DATE: 01-Jan-1996
; ATTORNEY/AGENT INFORMATION:
;   NAME: Williams, Ph.D., Kathleen M.
;   REGISTRATION NUMBER: 34,380
;   REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 345-9100
;   TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 707 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
US-08-726-306A-22

Query Match   8.3%; Score 28.4; DB 2; Length 707;
Best Local Similarity 46.1%; Pred. No. 1.1;

```
Qy    222 ggatgaacctggagctttattaccagtagagaggggaattcctccaccctgcccaacctgc 281
          ||||||| ||||||   |||||| | | ||||||||||    ::::::::::::::::::
Db   1024 GAATTAATTCCGAGCTTGGCTCGCAGGTCGCGAGGGAGCTTGCGGATYYYYYYYYYYYYYY 1083
Qy    282 tttcctctcccatgctcattaacactgttgtagatgctcatttttgtt 333
          :::::::::::::::::::::::::YYYYYYYYYYYYYYYYYYYYYYYYY
Db   1084 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1135

RESULT 15
US-08-848-252-3
; Sequence 3, Application US/08848252
; Patent No. 5804177
; GENERAL INFORMATION:
;   APPLICANT: Humphries, Keith R.
;   TITLE OF INVENTION: METHOD OF USING CD24 AS A CELL MARKER
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Bereskin & Parr
;     STREET: 40 King Street West
;     CITY: Toronto
;     STATE: Ontario
;     COUNTRY: Canada
;     ZIP: M5H 3Y2
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/848,252
;     FILING DATE: 29-APR-1997
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/08/538,052
;     FILING DATE:
;     APPLICATION NUMBER: US 08/151,672
;     FILING DATE: 15-NOV-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: McDiarmid, Shona S.
;     REGISTRATION NUMBER: P-38,798
;     REFERENCE/DOCKET NUMBER: 3158-028
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (416) 364-7311
;     TELEFAX: (416) 361-1398
;     TELEX: 06-23115
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1800 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   ORIGINAL SOURCE:
;     ORGANISM: Mus musculus
;   IMMEDIATE SOURCE:
;     LIBRARY: Mouse
;     CLONE: M1/69-J11d Heat-Stable Antigen
;   FEATURE:
;     NAME/KEY: sig_peptide
;     LOCATION: 73..150
;   FEATURE:
;     NAME/KEY: mat_peptide
;     LOCATION: 151..300
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 73..303
US-08-848-252-3

Query Match             8.2%;  Score 28.2;  DB 1;  Length 1800;
Best Local Similarity  50.4%;  Pred. No. 2.4;
Matches   69;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;
```

```
Matches   95;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;
Qy     32 attgaggtggtttgcaacgacgtctaggaagaaagtccgcgttaagtgcaacaccgat 91
          |||  | || ||| ||  |||||  || ||||  ||||  | ||| | |||||  || |
Db      1 ATGCAGAGTCTTCGTGAAAACCTTACCGGCACGACCATCACCCTTGAGGTGGGAGCCCAGT 60
Qy     92 gacaccatcggggacttgaagaaactgatagcggcccaaactggcactcgttggaataag 151
          ||||||||  | || ||||   ||||  | ||||||||||| ||||||   ||  ||| ||
Db     61 GACACCATCGAAAATGTGAAGGCCAAGATCCAGATAAGGAAGGCATTCCCCCCGACCAG 120
Qy    152 atcgttcttaaaaagtggtacacgatttttaaggaccatgtatctctgggagattatgaa 211
          ||     || ||||| |  |  | ||  || |||||  ||| |||| ||| |||| || |
Db    121 CAGAGGCTCATCTTTTGCAGGCAGCAGCTGGAAGATGGCCGTACTCTTTCTGACTACAAC 180
Qy    212 atccacgatggatgaacctggagct 237
          |||||  | ||  ||||  ||  |
Db    181 ATCCAGAAGGAGTCGACCCTGCACCT 206

RESULT 14
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
;   APPLICANT: DORNER, F.
;   APPLICANT: SCHEIFLINGER, F.
;   APPLICANT: FALKNER, F. G.
;   TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;   NUMBER OF SEQUENCES: 52
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Foley & Lardner
;     STREET: 1800 Diagonal Road, Suite 500
;     CITY: Alexandria
;     STATE: VA
;     COUNTRY: USA
;     ZIP: 22313-0299
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/232,463
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/07/935,313
;     FILING DATE:
;     APPLICATION NUMBER: EP 91 114 300.6
;     FILING DATE: 26-AUG-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME: BENT, Stephen A.
;     REGISTRATION NUMBER: 29,768
;     REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (703)836-9300
;     TELEFAX: (703)683-4109
;     TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 7218 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match             8.3%;  Score 28.4;  DB 1;  Length 7218;
Best Local Similarity  20.5%;  Pred. No. 4.6;
Matches   23;  Conservative   49;  Mismatches   49;  Indels   40;  Gaps    0;
```

```
Qy    204  attatgaaatccacgatgggatgaacctggagcttattaccagtagagggaattcctc  263
           ||| ||     | ||  ||||| |  |||  ||  |  ||||  |||  ||||  |        |
Db   1209  ATTCTACAGGCAAATATGCAAAGGAGCCAAAACTGTAAACCAGCATTTGGGATGTGAAG 1268

Qy    264  caccttgccaacctgctttcctcccatgctcatttaacactgttgtagatgctca  323
              | |||  |||  |  | ||  |  || | |||  || |   |||| || ||
Db   1269  ACTGGAAGCTAACTCTCATTGAATTCACAAGTCTTTATACAATTTCTGTACATACTTT 1328

Qy    324  tttttttgttaagtgta  340
           |||||||  ||||| | |
Db   1329  TTTTTTTTTAAGAGAA  1345
```

Search completed: June 13, 2001, 20:21:14
Job time: 6117 sec

OM nucleic - nucleic search, using sw model

```
              GenCore version 4.5
       Copyright (c) 1993 - 2000 Compugen Ltd.
```

Run on: June 13, 2001, 18:35:57 ; Search time 1224.15 Seconds
                       (without alignments)
                       4120.220 Million cell updates/sec

Title: US-09-331-930A-1
Perfect score: 342
Sequence: 1 gttccaggagattacagctc........atttttttgttaagtgtact 342

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

rched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*

1: gb_ba1:*	44: em_ov:*
2: gb_ba2:*	45: em_pat:*
3: gb_ba3:*	46: em_ph:*
4: gb_in1:*	47: em_pl:*
5: gb_in2:*	48: em_ro:*
6: gb_in3:*	49: em_sts:*
7: gb_om:*	50: em_sy:*
8: gb_ov:*	51: em_un:*
9: gb_pat1:*	52: em_vi:*
10: gb_pat2:*	53: gb_sts1:*
11: gb_ph:*	54: gb_sts2:*
12: gb_pl1:*	55: gb_sts3:*
13: gb_pl2:*	56: gb_sy:*
14: gb_pl3:*	57: gb_un:*
15: gb_pl4:*	58: gb_vi1:*
16: em_ba1:*	59: gb_vi2:*
17: em_ba2:*	60: gb_htg1:*
18: em_fun:*	61: gb_htg2:*
19: em_htgo_hum:*	62: gb_htg3:*
20: em_htgo_inv:*	63: gb_htg4:*
21: em_htgo_rod:*	64: gb_htg5:*
22: em_htg_hum1:*	65: gb_htg6:*
23: em_htg_hum2:*	66: gb_htg7:*
24: em_htg_hum3:*	67: gb_htg8:*
25: em_htg_hum4:*	68: gb_htg9:*
26: em_htg_hum5:*	69: gb_htg10:*
27: em_htg_hum6:*	70: gb_htg11:*
28: em_htg_hum7:*	71: gb_htg12:*
29: em_htg_hum8:*	72: gb_htg13:*
30: em_htg_inv1:*	73: gb_htg14:*
31: em_htg_inv2:*	74: gb_htg15:*
32: em_htg_other:*	75: gb_htg16:*
33: em_htg_rod:*	76: gb_htg17:*
34: em_hum1:*	77: gb_htg18:*
35: em_hum2:*	78: gb_htg19:*
36: em_hum3:*	79: gb_htg20:*
37: em_hum4:*	80: gb_htg21:*
38: em_hum5:*	81: gb_htg22:*
39: em_hum6:*	82: gb_htg23:*
40: em_hum7:*	83: gb_htg24:*
41: em_in:*	84: gb_htg25:*
42: em_om:*	85: gb_pr1:*
43: em_or:*	86: gb_pr2:*
	87: gb_pr3:*
	88: gb_pr4:*
	89: gb_pr5:*
	90: gb_pr6:*
	91: gb_pr7:*
	92: gb_pr8:*
	93: gb_pr9:*
	94: gb_ro1:*
	95: gb_ro2:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|---|
|  | 1 | 324 | 94.7 | 368 | 94 | AF318186 | AF318186 Psammomys |
| c | 2 | 234.6 | 68.6 | 204394 | 94 | AC079644 | AC079644 Mus muscu |
|  | 3 | 232.4 | 68.0 | 204394 | 94 | AC079644 | AC079644 Mus muscu |
|  | 4 | 195.2 | 57.1 | 176560 | 67 | AC021852 | AC021852 Homo sapi |
|  | 5 | 195.2 | 57.1 | 185380 | 64 | AC015941 | AC015941 Homo sapi |
|  | 6 | 195.2 | 57.1 | 186809 | 70 | AC026620 | AC026620 Homo sapi |
| c | 7 | 125.8 | 36.8 | 201602 | 75 | AC074146 | AC074146 Mus muscu |
| c | 8 | 125.8 | 36.8 | 219217 | 75 | AC069459 | AC069459 Mus muscu |
| c | 9 | 121.4 | 35.5 | 156342 | 70 | AC026142 | AC026142 Homo sapi |
| c | 10 | 121.4 | 35.5 | 160426 | 74 | AC068372 | AC068372 Homo sapi |
| c | 11 | 118.8 | 34.7 | 59762 | 12 | AB023032 | AB023032 Arabidops |

```
c  12  111  32.5   88010   13  ATT14D3              AL138649 Arabidops
c  13  111  32.5  182320   65  AC016808             AC016808 Homo sapi
c  14  109  31.9   94838   74  AC068782_4           Continuation (5 of
c  15  108  31.6     439       FHU10292             U10292 Fasciola he
c  16  102.8 30.1   1805   15  YSCPPGCS             M94269 Saccharomyc
c  17  102.8 30.1   2201   14  SCYNR032W            Z71647 S.cerevisia
c  18  102.8 30.1   3204    6  SCYNR033W            Z71648 S.cerevisia
c  19  102.8 30.1   3840   15  YSCPAMIBEN           L15299 Yeast para-
c  20   97.8 28.6   1065   53  CNS06PUK             AL409842 T3 end of
c  21   88.4 25.8   2906    6  DVAJ5174             AJ005174 Drosophil
c  22   78.8 23.0  236692   75  AC073773             AC073773 Mus muscu
c  23   78.8 23.0  266229   75  AC073676             AC073676 Mus muscu
c  24   75.6 22.1  139376   93  HSU95742             U95742 Human chrom
c  25   75.5 22.1  172827   86  AC007216             AC007216 Homo sapi
c  26   75   21.9   26412   66  AC020443             AC020443 Drosophil
c  27   75   21.9  111455   60  AC007549             AC007549 Drosophil
c  28   75   21.9  125593   60  AC008340             AC008340 Drosophil
c  29   70   20.5     871   54  CNS07D9X             AL440203 T3 end of
c  30   65.8 19.2   64036   72  AC040921             AC040921 Homo sapi
c  31   65.8 19.2  138036   87  AC008752             AC008752 Homo sapi
c  32   63.8 18.7   36532    5  CELF46F11            U88173 Caenorhabdi
c  33   52.8 15.4  192929   60  AC005505             AC005505 Plasmodiu
c  34   52.2 15.3   16920   14  SPBC31E1             AL049190 S.pombe c
c  35   52.2 15.3   17855   12  AC004698             AC004698 PGAK exte
c  36   49.2 14.4   65675   71  AC026985             AC026985 Homo sapi
c  37   43.4 12.7     932   53  CNS074H8             AL428802 clone BA0
c  38   40.8 11.9    3802   92  HSGEBCMA             Z29574 Homo sapien
c  39   38.4 11.2    1048   53  CNS076A0             AL431134 T7 end of
c  40   37.4 10.9  143172   70  AC026462             AC026462 Homo sapi
c  41   37    10.8     525    4  ACUBIFUS             X75628 A.castellan
c  42   36.8 10.7    1187   14  SCFPOLY              L41658 Saccharum h
c  43   36.4 10.7  173154   77  AC079933             AC079933 Trypanoso
c  44   36.4 10.6    1381   15  ZMU29158             U29158 Zea mays po
c  45   35.8 10.5    3531   14  SCU44026             U44026 Saccharomyc
```

ALIGNMENTS

RESULT 1
AF318186
LOCUS   AF318186   368 bp   mRNA   ROD   27-NOV-2000
DEFINITION Psammomys obesus beacon mRNA, complete cds.
ACCESSION   AF318186
VERSION   AF318186.1 GI:11345429
KEYWORDS   .
SOURCE   fat sand rat.
  ORGANISM Psammomys obesus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
            Psammomys.
REFERENCE   1 (bases 1 to 368)
  AUTHORS   Collier,G.R., McMillan,J.S., Windmill,K., Walder,K.,
            Tenne-Brown,J., de Silva,A., Trevaskis,J., Jones,S., Morton,G.J.,
            Lee,S., Augert,G., Civitarese,A. and Zimmet,P.Z.
  TITLE   Beacon: a novel gene involved in the regulation of energy balance
  JOURNAL   Diabetes 49 (11), 1766-1771 (2000)
  MEDLINE   20527879
   PUBMED   11078442
REFERENCE   2 (bases 1 to 368)
  AUTHORS   Collier,G.R., McMillan,J.S., Windmill,K., Walder,K.,
            Tenne-Brown,J., de Silva,A., Trevaskis,J., Jones,S., Morton,G.J.,
            Lee,S., Augert,G., Civitarese,A. and Zimmet,P.Z.
  TITLE   Direct Submission
  JOURNAL   Submitted (30-OCT-2000) Metabolic Research Unit, Deakin University,
            Pigdons Rd, Geelong, Vic 3217, Australia
FEATURES   Location/Qualifiers
     source   1..368
                     /organism="Psammomys obesus"
                     /db_xref="taxon:48139"
                     /tissue_type="hypothalamus"
                     /note="involved in control of energy balance"
     CDS   25..246

                     /codon_start=1
                     /product="beacon"
                     /protein_id="AAG34704.1"
                     /db_xref="GI:11345430"
                     /translation="MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQTGRWNKIVL
                     KKWYTIFKDHVSLGDYEIHDGMNLELYYQ"
     polyA_signal   115   343
BASE COUNT   115 a   78 g   81 c   94 t
ORIGIN   338. .343

Query Match   94.7%; Score 324; DB 94; Length 368;
Best Local Similarity   100.0%; Pred. No. 6e-95;
Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   5 caggagattacagctccagccacaatgattgagctggtttgcaacgaccgtctaggaaag 64
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CAGGAGATTACAGCTCCAGCCACAATGATTGAGCTGGTTTGCAACGACCGTCTAGGAAAG 60

Qy   65 aaagtccgcgttaagtgcaacaccgatgacaccatcgggactgaagaaactgatagcg 124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAAGTCCGCGTTAAGTGCAACACCGATGACACCATCGGGACTGAAGAAACTGATAGCG 120

Qy   125 gcccaaactggcactcgttggaataagatcgttcttaaaaagtgtacacgatttttaag 184
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCCCAAACTGGCACTCGTTGGAATAAGATCGTTCTTAAAAAGTGTACACGATTTTTAAG 180

Qy   185 gaccatgtatctctgggagattatgaaatccacgatggggatgaacctgagctttattac 244
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GACCATGTATCTCTGGGAGATTATGAAATCCACGATGGGGATGAACCTGAGCTTTATTAC 240

Qy   245 cagtagagggaattcctccacttgccaacttgctttcctcctcccatggctcattta 304
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CAGTAGAGGGAATTCCTCCACCTTGCCCAACCTTGCTTTCCTCCTCCCATGGCTCATTTA 300

Qy   305 acactgttgtagatgctcattttt 328
          ||||||||||||||||||||||
Db   301 ACACTGTTGTAGATGCTCATTTTT 324

RESULT 2
AC079644/c
LOCUS   AC079644   204394 bp   DNA   ROD   01-JAN-2001
DEFINITION Mus musculus 12 BAC RP23-17406 (Roswell Park Cancer Institute Mouse
            BAC Library) complete sequence.
ACCESSION   AC079644
VERSION   AC079644.9 GI:12000470
KEYWORDS   HTG.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 204394)
  AUTHORS   Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
            Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
            Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
            Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
            Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
            Fernandez,C., Ferraguto,P., Forcum-Tansey,J., Gill,R.,
            Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
            Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
            Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
            Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
            Neal,D., Nelson,A., Nguyen,N., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
            Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
            Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
            Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
            Worley,K. and Gibbs,R.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 204394)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission

```
JOURNAL     Submitted (07-SEP-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 204394)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (01-JAN-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 1, 2001 this sequence version replaced gi:11991738.
            Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.

            Exon/Intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.
FEATURES             Location/Qualifiers
     source          1..204394
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="12"
                     /clone="RP23-174O6"
     repeat_region   2..273
                     /rpt_family="L1"
     repeat_region   complement(272..912)
                     /rpt_family="Lx2"
     repeat_region   1019..1138
                     /rpt_family="B1_MM"
     repeat_region   1243..1336
                     /rpt_family="B3"
     repeat_region   complement(1506..1575)
                     /rpt_family="PB1"
     repeat_region   2079..2206
                     /rpt_family="PB1D10"
     repeat_region   2222..2287
                     /rpt_family="RSINE1"
     repeat_region   2288..2310
                     /rpt_family="(TTAAA)n"
     repeat_region   2411..2613
                     /rpt_family="B3"
     repeat_region   3061..3143
                     /rpt_family="T-rich"
     repeat_region   complement(3144..3284)
                     /rpt_family="B1_MM"
     repeat_region   3285..3302
                     /rpt_family="T-rich"
     repeat_region   3316..3443
                     /rpt_family="B3"
     repeat_region   3584..3690
                     /rpt_family="B1-F"
     repeat_region   3691..3735
                     /rpt_family="B1_MM"
     repeat_region   3736..3756
                     /rpt_family="(CA)n"
     repeat_region   4008..4163
                     /rpt_family="B1_MM"
     repeat_region   5394..5823
                     /rpt_family="URR1A"
     repeat_region   5851..6040
                     /rpt_family="MT-INTERNAL"
     repeat_region   6946..7032
                     /rpt_family="B3A"
     repeat_region   7475..7769
                     /rpt_family="RLTRETN_MM"
     repeat_region   7782..7805
                     /rpt_family="(CAGA)n"
     repeat_region   complement(8113..8211)
                     /rpt_family="ORR1A2"
     repeat_region   8711..8779
                     /rpt_family="(TTTTTG)n"
     repeat_region   complement(8839..9415)
                     /rpt_family="L1"
     repeat_region   9419..9656
                     /rpt_family="Lx2B"
     repeat_region   10068..10256
                     /rpt_family="B2"
     repeat_region   10355..10506
                     /rpt_family="(TA)n"
     repeat_region   10690..10728
                     /rpt_family="(CAAAA)n"
     misc_feature    11537..11693
                     /rpt_family="B1_MM"
     misc_feature    complement(14141..14699)
                     /note="Region similar to Mm#S355374 AU015316 Mus musculus
                     cDNA:gb=AU015316"
     repeat_region   15017..16339
                     /rpt_family="L1F"
     repeat_region   complement(20743..21668)
                     /rpt_family="ORR1A-INT"
     repeat_region   complement(21669..21990)
                     /rpt_family="ORR1A2"
     repeat_region   complement(22241..22307)
                     /rpt_family="B4"
     repeat_region   22609..22758
                     /rpt_family="B3"
     repeat_region   complement(22809..22937)
                     /rpt_family="RSINE1"
     repeat_region   complement(23153..23297)
                     /rpt_family="B1_MM"
     repeat_region   complement(23359..23413)
                     /rpt_family="B2"
     repeat_region   complement(23900..24050)
                     /rpt_family="B1_MM"
     repeat_region   complement(24051..24168)
                     /rpt_family="B2"
     repeat_region   complement(24183..24350)
                     /rpt_family="B2"
     repeat_region   complement(24491..24571)
                     /rpt_family="B1_MM"
     repeat_region   24607..24765
                     /rpt_family="ORR1A3"
     repeat_region   24770..24876
                     /rpt_family="B1_MM"
     repeat_region   25119..25262
                     /rpt_family="B2"
     repeat_region   25306..25427
                     /rpt_family="(TTCC)n"
     repeat_region   25741..25813
                     /rpt_family="MTD"
     repeat_region   25814..25962
                     /rpt_family="B3"
     repeat_region   complement(26229..26333)
                     /rpt_family="Lx7"
     repeat_region   complement(26553..26756)
                     /rpt_family="B2"
     repeat_region   26842..27153
                     /rpt_family="B2"
     misc_feature    27339..27790
                     /note="Region similar to Mm#S189632 vd29g11.s1 Mus
                     musculus cDNA:gb=AA422861"
     gene            complement(join(27558..27790,34041..34255,65878..66313,
                     72368..72803,81034..81465,137727..138158,193391..193825))
                     /gene="gi|2580559|gb|AF000987|HSAF000987 Homo sapiens
                     eIF-1A, Y isoform (EIF1AY) mRNA"
     repeat_region   complement(35572..35704)
                     /rpt_family="B1_MM"
     repeat_region   complement(35705..35802)
```

```
                    Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
                    Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
                    Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
                    Neal,D., Nelson,A., Nguyen,N., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
                    Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
                    Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
                    Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
                    Worley,K. and Gibbs,R.
  TITLE      Direct Submission
  JOURNAL    Unpublished
REFERENCE    2 (bases 1 to 204394)
  AUTHORS    Worley,K.C.
  TITLE      Direct Submission
  JOURNAL    Submitted (07-SEP-2000) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE    3 (bases 1 to 204394)
  AUTHORS    Worley,K.C.
  TITLE      Direct Submission
  JOURNAL    Submitted (01-JAN-2001) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Jan 1, 2001 this sequence version replaced gi:11991738.
             Sequencing is completed to a minimum standard of double strand
             coverage with a minimum of 2 clones and 2 reads with no ambiguities
             or 2 chemistries with a minimum of 2 clones and 3 reads with no
             ambiguities. If the sequence quality does not meet this standard,
             it will be indicated in the annotation.

             The repeat regions shown were identified using RepeatMasker by
             Adrian Smit.

             Sequence similarities were identified using Powerblast by Jinghui
             Zhang.

             Exon/Intron boundaries of identified genes were chosen if there
             were canonical splice junctions that maintained sequence continuity
             across the splice junctions.
FEATURES             Location/Qualifiers
     source          1..204394
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="12"
                     /clone="RP23-17406"
     repeat_region   2..273
                     /rpt_family="L1"
     repeat_region   complement(272..912)
                     /rpt_family="Lx2"
     repeat_region   1019..1138
                     /rpt_family="B1_MM"
     repeat_region   1243..1336
                     /rpt_family="B3"
     repeat_region   complement(1506..1575)
                     /rpt_family="PB1"
     repeat_region   2079..2206
                     /rpt_family="PB1D10"
     repeat_region   2222..2287
                     /rpt_family="RSINE1"
     repeat_region   2288..2310
                     /rpt_family="(TTAAA)n"
     repeat_region   2411..2613
                     /rpt_family="B3"
     repeat_region   3061..3143
                     /rpt_family="T-rich"
     repeat_region   complement(3144..3284)
                     /rpt_family="B1_MM"
     repeat_region   3285..3302
                     /rpt_family="T-rich"
     repeat_region   3316..3443
                     /rpt_family="B1-F"
     repeat_region   3584..3690
                     /rpt_family="B1_MM"
     repeat_region   3691..3735
```

```
     repeat_region   /rpt_family="B2"
                     complement(35806..35931)
                     /rpt_family="Lx2"
                     complement(35930..36646)
                     /rpt_family="Lx2"
     repeat_region   complement(36961..37371)
                     /rpt_family="MULV"
     repeat_region   39203..39223
                     /rpt_family="RMER2"
     repeat_region   39224..39617
                     /rpt_family="MTA"
     repeat_region   39618..39728
                     /rpt_family="RMER2"
     repeat_region   39728..40572
                     /rpt_family="ORR1A-INT"
     repeat_region   40582..40880
                     /rpt_family="ORR1A2"
     repeat_region   complement(41260..41315)
                     /rpt_family="ORR1B"
     repeat_region   complement(41320..41920)
                     /rpt_family="L1_MM"
     repeat_region   complement(41921..42123)
                     /rpt_family="L1"

Query Match    68.68;   Score 234.6;  DB 94;  Length 204394;
Best Local Similarity   85.78;   Pred. No. 3.2e-65;
Matches  287;  Conservative  0;  Mismatches  39;  Indels  9;  Gaps  2;

Qy      2 ttccaggagattacagctccagccacacaatgattgagtggttgcaacgaccgtctagga 61
          ||| ||||   |||||||||| |||||||||||||  |||||| |||||||||||||||||
Db  49786 TTCGAGGACATTGGAGCTCCAGCCACCAATGATTGAGGTGGTTTGCAATGACCGTCTAGGA 49727

Qy     62 aagaaagtccgttaagtgcaacaccgatgacacaccgcgggagacttgaagaaactgata 121
          |||||||| || ||||| || ||||||||||||||||| |||| | ||||||||||||||
Db  49726 AAGAAAGTCCGAGTGAAGTGCAACACCGATGATACCATTGGTGGCTTGAAGAAACTGATA 49667

Qy    122 gcgcccaaactggcactcgttggaataagatcgttcttaaaaagtgtaccagatttttt 181
          |||||| ||||| || |||| || |||| |||| |||| |||||||||| ||||||||
Db  49666 GCGGCTCAAACTGGCACCCGCCGTTGAATAAGATCATTCTTAAAAAGTGGTACACGATTTAT 49607

Qy    182 aaggaccatgtatctctgggagattatgaaatccacgatggatgaacctggagctttat 241
          ||||||| |||||||||||||| |||||||||||||||||||||| ||||| || |||||
Db  49606 AAGGACCACCGTGTCTCTGGGAGACTATGAAATCCATGATGGGATGAACCTGGAGCTTTAT 49547

Qy    242 taccagtagagggaattcctccacct------tgcccaacctttgctttcctctccca 293
          ||||||||||| |||||||| ||||||      | ||| ||||| ||||| |||||
Db  49546 TACCAGTAGAGCGGAATTCTTCCTCCTCTTGCCCCCACTCTACCCTGCCTTCCTATCCCA 49487

Qy    294 tggctcattaacactgttgtagatgctcatttt 328
          |||||||||| || ||||| || |||| |||||||
Db  49486 -CCCTCATTTGAAACTTCTATAGATGGTCATTTTT 49453

RESULT  3
AC079644
LOCUS       AC079644   204394 bp   DNA   ROD   01-JAN-2001
DEFINITION  Mus musculus 12 BAC RP23-17406 (Roswell Park Cancer Institute Mouse
            BAC Library) complete sequence.
ACCESSION   AC079644
VERSION     AC079644.9   GI:12000470
KEYWORDS    HTG.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 204394)
  AUTHORS   Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
            Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
            Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
            Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
            Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
            Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
            Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
```

```
repeat_region      /rpt_family="(CA)n"
                   3736..3756
repeat_region      /rpt_family="B1_MM"
                   4008..4163
repeat_region      /rpt_family="URR1A"
                   5394..5823
repeat_region      /rpt_family="MT-INTERNAL"
                   5851..6040
repeat_region      /rpt_family="B3A"
                   6946..7032
repeat_region      /rpt_family="B1_MM"
                   7475..7769
repeat_region      /rpt_family="RLTRETN_MM"
                   7782..7805
repeat_region      /rpt_family="(CAGA)n"
                   complement(8113..8211)
repeat_region      /rpt_family="ORR1A2"
                   8711..8779
repeat_region      /rpt_family="(TTTTTG)n"
                   complement(8839..9415)
repeat_region      /rpt_family="L1"
                   9419..9656
repeat_region      /rpt_family="Lx2B"
                   10068..10256
repeat_region      /rpt_family="B2"
                   10355..10506
repeat_region      /rpt_family="(TA)n"
                   10690..10728
repeat_region      /rpt_family="(CAAAA)n"
                   11537..11693
misc_feature       /rpt_family="B1_MM"
                   complement(14141..14699)
                   /note="Region similar to Mm#S355374 AU015316 Mus musculus
                   cDNA:gb=AU015316"
                   15017..16339
repeat_region      /rpt_family="L1F"
                   complement(20743..21668)
repeat_region      /rpt_family="ORR1A-INT"
                   complement(21669..21990)
repeat_region      /rpt_family="ORR1A2"
                   complement(22241..22307)
repeat_region      /rpt_family="B4"
                   22609..22758
repeat_region      /rpt_family="B3"
                   complement(22809..22937)
repeat_region      /rpt_family="RSINE1"
                   complement(23153..23297)
repeat_region      /rpt_family="B1_MM"
                   complement(23359..23413)
repeat_region      /rpt_family="B2"
                   complement(23900..24050)
repeat_region      /rpt_family="B1_MM"
                   complement(24051..24168)
repeat_region      /rpt_family="B2"
                   complement(24183..24350)
repeat_region      /rpt_family="B2"
                   complement(24491..24571)
repeat_region      /rpt_family="B1_MM"
                   24607..24765
repeat_region      /rpt_family="ORR1A3"
                   24770..24876
repeat_region      /rpt_family="B1_MM"
                   25119..25262
repeat_region      /rpt_family="(TTCC)n"
                   25306..25427
repeat_region      /rpt_family="MTD"
                   25741..25813
repeat_region      /rpt_family="B3"
                   25814..25962
repeat_region      /rpt_family="Lx7"
                   complement(26229..26333)
repeat_region      /rpt_family="B2"
                   complement(26553..26756)
```

```
repeat_region      /rpt_family="B2"
                   26842..27153
repeat_region      /rpt_family="Lx2"
                   27339..27790
misc_feature       /note="Region similar to Mm#S189632 vd29g11.s1 Mus
                   musculus cDNA:gb=AA422861"
gene               complement(join(27558..27790,34041..34255,65878..66313,
                   72368..72803,81034..81465,137727..138158,193391..193825))
                   /gene="gi|2580559|gb|AF000987|HSAF000987 Homo sapiens
                   eIF-1A, Y isoform (EIF1AY) mRNA"
repeat_region      complement(35572..35704)
                   /rpt_family="B1_MM"
repeat_region      complement(35705..35802)
                   /rpt_family="B2"
repeat_region      complement(35806..35931)
                   /rpt_family="Lx2"
repeat_region      complement(35930..36646)
                   /rpt_family="ORR1A2"
repeat_region      complement(36961..37371)
                   /rpt_family="Lx2"
repeat_region      39203..39223
                   /rpt_family="MULV"
repeat_region      39224..39617
                   /rpt_family="RMER2"
repeat_region      39618..39728
                   /rpt_family="MTA"
repeat_region      39728..40572
                   /rpt_family="RMER2"
repeat_region      40582..40880
                   /rpt_family="ORR1A-INT"
repeat_region      complement(41260..41315)
                   /rpt_family="ORR1A2"
repeat_region      complement(41320..41920)
                   /rpt_family="ORR1B"
repeat_region      complement(41921..42123)
                   /rpt_family="L1_MM"
repeat_region      /rpt_family="L1"
```

```
Query Match       Best Local Similarity   68.0%;   Score 232.4;   DB 94;   Length 204394;
                  85.1%;   Pred. No. 1.7e-64;
Matches 286;   Conservative   0;   Mismatches   41;   Indels   9;   Gaps   2;

Qy        1 gttccaggagattacagctccagccacacaatgattgagtggttgcaacgaccgtctagg   60
            ||| |||||||||||||||||||||||||||||||| |||||||||||||| |||| ||||
Db   112084 GTTCGAGGAGACATTGGAGCTCCAGCCACAATGATTGAGTGGTTGCAATGACCATCTAGG 112143

Qy       61 aaagaaagtccgcgttaagtgcaacaccgatgacaccatcggggacttgaagaaactgat  120
            |||||||| || | ||||||||||||||||||| ||||||||||||||||||||||||||
Db   112144 AAAGAAAGTCTGAGTGAAGTGCAACACCATCGATGACACTACCATTGGTGACTTGAAGAAACTGAT 112203

Qy      121 agcggccaaactggcactcgttggaataagatcgtcttaaaaagtggtacacgatttt  180
            ||| || |||||||||| | ||||||||||||| ||  ||||||||| |||||||| |
Db   112204 AGCCGCTCAAACTGGCACCCGCTGGAATAAGATCATTCTTAAAAAGTGTACACGATTTA 112263

Qy      181 taaggaccatgtatctctgggagattagaaatccacgatgggatgacatggatggagcttta  240
            ||||||||| || ||||| ||||| || ||||||| || ||||| || || ||||||||||
Db   112264 TAAGGACCACCGTGTCTCTCGGAGACTATGAAATCCATGAAATCCATGATAGGATGAACCTGAGCTTTA 112323

Qy      241 ttaccagtagaggggaattcctcacct------tgcccaaccttgcttcctctccc  292
            ||||||||||| | ||||| || ||       ||||| || | ||| ||| | ||
Db   112324 TTACCAGTAGAATGGAATTCTTCCTCCCTCCTTGCCCCACTCACCCTGCCTATCCC 112383

Qy      293 atggctcattaacactgtttagatgctcattttt  328
            |  ||||||| ||||| | |||||||| ||||||
Db   112384 A-CCCTCATTTGACACTTCTATAGATGTCATTTTT 112418
```

```
RESULT    4
AC021852
LOCUS       AC021852   176560 bp   DNA   HTG   07-JUL-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-474K4, WORKING DRAFT
            SEQUENCE, 30 unordered pieces.
ACCESSION   AC021852
VERSION     AC021852.4   GI:8868867
```

```
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176560)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176560)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 16, 2000 this sequence version replaced gi:7263413.
            ----------- Genome Center -----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ----------- Project Information -----------
            Center project name: H_NH0474K04
            ----------- Summary Statistics -----------
            Sequencing vector: M13; 74%
            Sequencing vector: plasmid; 26%
            Chemistry: Dye-primer ET; 74% of reads
            Chemistry: Dye-terminator Big Dye; 26% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 161959 bases at least Q40
            Consensus quality: 165336 bases at least Q30
            Consensus quality: 167883 bases at least Q20
            Insert size: 168000; agarose-fp
            Insert size: 173660; sum-of-contigs
            Quality coverage: 3.87 in Q20 bases; agarose-fp
            Quality coverage: 3.82 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 30 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1520: contig of 1520 bp in length
            *     1521     1620: gap of unknown length
            *     1621     3161: contig of 1541 bp in length
            *     3162     3261: gap of unknown length
            *     3262     5461: contig of 2200 bp in length
            *     5462     5561: gap of unknown length
            *     5562     7149: contig of 1588 bp in length
            *     7150     7249: gap of unknown length
            *     7250    11327: contig of 4078 bp in length
            *    11328    11427: gap of unknown length
            *    11428    13630: contig of 2203 bp in length
            *    13631    13730: gap of unknown length
            *    13731    15908: contig of 2178 bp in length
            *    15909    16008: gap of unknown length
            *    16009    18540: contig of 2532 bp in length
            *    18541    18640: gap of unknown length
            *    18641    20475: contig of 1835 bp in length
            *    20476    20575: gap of unknown length
            *    20576    22994: contig of 2419 bp in length
            *    22995    23094: gap of unknown length
            *    23095    26651: contig of 3557 bp in length
            *    26652    26751: gap of unknown length
            *    26752    31473: contig of 4722 bp in length
            *    31474    31573: gap of unknown length
            *    31574    35428: contig of 3855 bp in length
            *    35429    35528: gap of unknown length
            *    35529    40087: contig of 4559 bp in length
            *    40088    40187: gap of unknown length
            *    40188    44325: contig of 4138 bp in length
            *    44326    44425: gap of unknown length
            *    44426    49492: contig of 5067 bp in length
            *    49493    49592: gap of unknown length
            *    49593    57072: contig of 7480 bp in length
            *    57073    57172: gap of unknown length
            *    57173    63683: contig of 6511 bp in length
            *    63684    63783: gap of unknown length
            *    63784    70585: contig of 6802 bp in length
            *    70586    70685: gap of unknown length
            *    70686    76995: contig of 6310 bp in length
            *    76996    77095: gap of unknown length
            *    77096    82834: contig of 5739 bp in length
            *    82835    82934: gap of unknown length
            *    82935    89274: contig of 6340 bp in length
            *    89275    89374: gap of unknown length
            *    89375    97794: contig of 8420 bp in length
            *    97795    97894: gap of unknown length
            *    97895   107952: contig of 10058 bp in length
            *   107953   108052: gap of unknown length
            *   108053   117316: contig of 9264 bp in length
            *   117317   117416: gap of unknown length
            *   117417   128399: contig of 10083 bp in length
            *   128400   128499: gap of unknown length
            *   128500   139936: contig of 11437 bp in length
            *   139937   140036: gap of unknown length
            *   140037   149771: contig of 9735 bp in length
            *   149772   149871: gap of unknown length
            *   149872   163648: contig of 13777 bp in length
            *   163649   163748: gap of unknown length
            *   163749   176560: contig of 12812 bp in length.
FEATURES             Location/Qualifiers
     source          1..176560
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-474K4"
     misc_feature    1..1520
                     /note="assembly_name:Contig6"
     misc_feature    1621..3161
                     /note="assembly_name:Contig7"
     misc_feature    3262..5461
                     /note="assembly_name:Contig8"
     misc_feature    5562..7149
                     /note="assembly_name:Contig9"
     misc_feature    7250..11327
                     /note="assembly_name:Contig10"
     misc_feature    11428..13630
                     /note="assembly_name:Contig11"
     misc_feature    13731..15908
                     /note="assembly_name:Contig12"
     misc_feature    16009..18540
                     /note="assembly_name:Contig13"
     misc_feature    18641..20475
                     /note="assembly_name:Contig14"
                     clone_end:T7
                     vector_side:right"
     misc_feature    20576..22994
                     /note="assembly_name:Contig15"
     misc_feature    23095..26651
                     /note="assembly_name:Contig16"
     misc_feature    26752..31473
                     /note="assembly_name:Contig17"
     misc_feature    31574..35428
                     /note="assembly_name:Contig18"
     misc_feature    35529..40087
                     /note="assembly_name:Contig19"
     misc_feature    40188..44325
                     /note="assembly_name:Contig20"
     misc_feature    44426..49492
                     /note="assembly_name:Contig21"
     misc_feature    49593..57072
                     /note="assembly_name:Contig22"
```

```
misc_feature    57173. .63683
                /note="assembly_name:Contig23"
misc_feature    63784. .70585
                /note="assembly_name:Contig24"
misc_feature    70686. .76995
                /note="assembly_name:Contig25"
misc_feature    77096. .82834
                /note="assembly_name:Contig26"
misc_feature    82935. .89274
                /note="assembly_name:Contig27
                clone_end:SP6
                vector_side:right"
misc_feature    89375. .97794
                /note="assembly_name:Contig28"
misc_feature    97895. .107952
                /note="assembly_name:Contig29"
misc_feature    108053. .117316
                /note="assembly_name:Contig30"
misc_feature    117417. .128399
                /note="assembly_name:Contig31"
misc_feature    128500. .139936
                /note="assembly_name:Contig32"
misc_feature    140037. .149771
                /note="assembly_name:Contig33"
misc_feature    149872. .163648
                /note="assembly_name:Contig34"
misc_feature    163749. .176560
                /note="assembly_name:Contig35"
BASE COUNT    50382 a    37345 c    38048 g    47863 t    2922 others
ORIGIN
```

```
REFERENCE      1 (bases 1 to 185380)
AUTHORS        Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE          Homo sapiens chromosome, clone RP11-55J8
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 185380)
AUTHORS        Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
               Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
               Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
               Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
               Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.
               Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
               Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
               Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
               McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
               Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
               Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
               Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
               Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
               Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE          Direct Submission
JOURNAL        Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
               Research, 320 Charles Street, Cambridge, MA 02141, USA
               On Apr 1, 2000 this sequence version replaced gi:6939304.
               All repeats were identified using RepeatMasker:
               Smit, A.F.A. & Green, P. (1996-1997)
               http://ftp.genome.washington.edu/RM/RepeatMasker.html
COMMENT        ---------------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
               ---------------- Project Information
               Center project name: L920
               Center clone name: 55_J_8
               ---------------- Summary Statistics
               Sequencing vector: M13; M77815; 100% of reads
               Chemistry: Dye-primer-amersham; 3% of reads
               Chemistry: Dye-terminator Big Dye; 97% of reads
               Assembly program: Phrap; version 0.960731
               Consensus quality: 171625 bases at least Q40
               Consensus quality: 177767 bases at least Q30
               Consensus quality: 180190 bases at least Q20
               Insert size: 189000; agarose-fp
               Insert size: 182080; sum-of-contigs
               Quality coverage: 4.8 in Q20 bases; agarose-fp
               Quality coverage: 5.0 in Q20 bases;.
             * NOTE:This is a 'working draft' sequence. It currently
             * consists of 34 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *        1     1291: contig of 1291 bp in length
             *  1292  1391: gap of      100 bp
             *  1392  3276: contig of 1885 bp in length
             *  3277  3376: gap of      100 bp
             *  3377  4744: contig of 1368 bp in length
             *  4745  4844: gap of      100 bp
             *  4845  6631: contig of 1787 bp in length
             *  6632  6731: gap of      100 bp
             *  6732  8626: contig of 1895 bp in length
             *  8627  8726: gap of      100 bp
             *  8727 10349: contig of 1623 bp in length
             * 10350 10449: gap of      100 bp
             * 10450 11784: contig of 1335 bp in length
             * 11785 11884: gap of      100 bp
             * 11885 14034: contig of 2150 bp in length
             * 14035 14134: gap of      100 bp
             * 14135 15586: contig of 1452 bp in length
             * 15587 15686: gap of      100 bp
             * 15687 17856: contig of 2170 bp in length
             *
```

```
Query Match           57.1%;    Score 195.2;  DB 67;   Length 176560;
Best Local Similarity 78.7%;    Pred. No. 2.5e-52;
Matches 258;  Conservative   0;  Mismatches  68;  Indels   2;  Gaps   2;

Qy         2 ttccagagattacagctccagccacaatgattgaggtggttgcaacgaccgtctagga 61
             | |-|||  ||| ||||||||| |  |||||  ||||||| ||||||||||||| ||
Db    174562 TGCAAGGCAATCGACCTCAGCTAGGATGATCCAGGTTGTTTGCAACCACCGTCTGGGG 174621

Qy        62 aagaaagtccgcgttaagtgcaacaccgatgacacacctgggacttgaagaaactgata 121
             || |-|||  ||||||| ||||||||||  |||||||  ||||||||||| ||||
Db    174622 AAAAAGGTCCACGTTAAATGCAACACCGATGATACCACCGGGGACCTTAAGAAGCTGATT 174681

Qy       122 gcggcccaactggcactcgttggaataagatcgtcttcttaaaaagtggtacacgatttt 181
             ||-||| ||||| ||  ||||| ||  ||-|| ||  |||| ||||||| |||||||||
Db    174682 GCAGCCTAAACTGGCACCCGTCGGAACACAAGATTGTCCTGAAGAAGTGGTACACAATTTTT 174741

Qy       182 aaggaccatgtatctctgggagattatgaaatccacgatgggatgaacctgagctttat 241
             ||||||||||| || |||||||||||  || |||||||||||||||||| ||||||||
Db    174742 AAGGACCATGTGTCTCTCGGGGACTATGAAATCCACAATGGGATGAACCTGGAGCTTTAT 174801

Qy       242 taccagtagagggaattcctccaaccttgcccaaccttgctttcctctcccat-ggctca 300
             ||||| | | ||   |||| ||| || |||||||||| ||| |||||||  |||||
Db    174802 TATCAAAAGATGAGAA-TCCTCATCTTCCTGCGTGCCCACCTCTCCCATCCTCACC 174860

Qy       301 tttaacactgttgtagatgctcatttt 328
             -|||| |||||||||||||| ||| |||
Db    174861 CCGCACACTGGGGATAGATGCTTGTTTGT 174888

RESULT 5
AC015941
LOCUS       AC015941    185380 bp    DNA                   HTG           01-APR-2000
DEFINITION  Homo sapiens clone RP11-55J8, WORKING DRAFT SEQUENCE, 34 unordered
            pieces.
ACCESSION   AC015941
VERSION     AC015941.5   GI:7382668
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
*        17857   17956: gap of     100 bp
*        17957   20816: contig of 2860 bp in length
*        20817   20916: gap of     100 bp
*        20917   23798: contig of 2882 bp in length
*        23799   23898: gap of     100 bp
*        23899   26035: contig of 2137 bp in length
*        26036   26135: gap of     100 bp
*        26136   28954: contig of 2819 bp in length
*        28955   29054: gap of     100 bp
*        29055   33642: contig of 4588 bp in length
*        33643   33742: gap of     100 bp
*        33743   36791: contig of 3049 bp in length
*        36792   36891: gap of     100 bp
*        36892   39359: contig of 2468 bp in length
*        39360   39459: gap of     100 bp
*        39460   44800: contig of 5341 bp in length
*        44801   44900: gap of     100 bp
*        44901   50038: contig of 5138 bp in length
*        50039   50138: gap of     100 bp
*        50139   53537: contig of 3399 bp in length
*        53538   53637: gap of     100 bp
*        53638   59913: contig of 6276 bp in length
*        59914   60013: gap of     100 bp
*        60014   65717: contig of 5704 bp in length
*        65718   65817: gap of     100 bp
*        65818   73361: contig of 7544 bp in length
*        73362   73461: gap of     100 bp
*        73462   80860: contig of 7399 bp in length
*        80861   80960: gap of     100 bp
*        80961   89233: contig of 8273 bp in length
*        89234   89333: gap of     100 bp
*        89334   96487: contig of 7154 bp in length
*        96488   96587: gap of     100 bp
*        96588  105683: contig of 9096 bp in length
*       105684  105783: gap of     100 bp
*       105784  113921: contig of 8138 bp in length
*       113922  114021: gap of     100 bp
*       114022  122911: contig of 8890 bp in length
*       122912  123011: gap of     100 bp
*       123012  134520: contig of 11509 bp in length
*       134521  134620: gap of     100 bp
*       134621  144215: contig of 9595 bp in length
*       144216  144315: gap of     100 bp
*       144316  154607: contig of 10292 bp in length
*       154608  154707: gap of     100 bp
*       154708  169066: contig of 14359 bp in length
*       169067  169166: gap of     100 bp
*       169167  185380: contig of 16214 bp in length.

FEATURES             Location/Qualifiers
     source          1..185380
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-55J8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1291
                     /note="assembly_fragment"
     misc_feature    1392..3276
                     /note="assembly_fragment"
     misc_feature    3377..4744
                     /note="assembly_fragment"
     misc_feature    4845..6631
                     /note="assembly_fragment"
     misc_feature    6732..8626
                     /note="assembly_fragment"
     misc_feature    8727..10349
                     /note="assembly_fragment"
     misc_feature    10450..11784
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    11885..14034
                     /note="assembly_fragment"
     misc_feature    14135..15586
```

```
     misc_feature    15687..17856
                     /note="assembly_fragment"
     misc_feature    17957..20816
                     /note="assembly_fragment"
     misc_feature    20917..23798
                     /note="assembly_fragment"
     misc_feature    23899..26035
                     /note="assembly_fragment"
     misc_feature    26136..28954
                     /note="assembly_fragment"
     misc_feature    29055..33642
                     /note="assembly_fragment"
     misc_feature    33743..36791
                     /note="assembly_fragment"
     misc_feature    36892..39359
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    39460..44800
                     /note="assembly_fragment"
     misc_feature    44901..50038
                     /note="assembly_fragment"
     misc_feature    50139..53537
                     /note="assembly_fragment"
     misc_feature    53638..59913
                     /note="assembly_fragment"
     misc_feature    60014..65717
                     /note="assembly_fragment"
     misc_feature    65818..73361
                     /note="assembly_fragment"
     misc_feature    73462..80860
                     /note="assembly_fragment"
     misc_feature    80961..89233
                     /note="assembly_fragment"
     misc_feature    89334..96487
                     /note="assembly_fragment"
     misc_feature    96588..105683
                     /note="assembly_fragment"
     misc_feature    105784..113921
                     /note="assembly_fragment"
     misc_feature    114022..122911
                     /note="assembly_fragment"
     misc_feature    123012..134520
                     /note="assembly_fragment"
```

```
Query Match             57.1%;  Score 195.2;  DB 64;  Length 185380;
Best Local Similarity   78.7%;  Pred. No. 2.5e-52;
Matches  258;  Conservative    0;  Mismatches   68;  Indels    2;  Gaps    2;

Qy      2 ttccaggagattacagctccagccacaatgattgagtggttcaacgaccgtctagga 61
          | ||| |||| ||| | ||||||| | |||| |||| ||||||||||||||||| ||
Db  24462 TGCAAGGCAATTCGAGCTCCAGCTAGGATGATCGAGGTTGTTCAACGACCGTCTGGGG 24521

Qy     62 aagaaagtccgcgttaagtgcaacaccgatgacaccatcgggggacttgaagaaactgata 121
          || || ||||||||| ||| |||||| ||||| ||||| |||| || |||| ||| ||||
Db  24522 AAAAAGGTCCACGTTAAATGCAACACGGATGATACCATCGGGGACCTTAAGAAGCTGATT 24581

Qy    122 gcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatttt 181
          || ||| ||||||||| |||| |||| || ||| ||| ||||| | |||||||| |||||
Db  24582 GCAGCCTAAACTGGCACCGTTGGAACAAGATTGTCCTGAACAAGATGGTACACAATTTTT 24641

Qy    182 aaggaccatgtatcctctgggagattagaaatccacgatgaaatccacggaggcttat 241
          ||||||||| | ||| ||||| || | ||||||||| ||||| |||| |||||||||||
Db  24642 AAGGACCATGTGTCTCTGGGGGACTATGAAATCCACAATGGATGAGAAATCCACGGAGCTTTAT 24701

Qy    242 taccagtagaggggaattcctccacctgccaacctttctcttccccat-ggctca 300
          ||| | ||||| |||||||| | | |||||||| | | |||||||| |  |||||
Db  24702 TATCAAAAGATGAGAA-TCCTCATCTTCCGCTGCCCCACTCTCCTCCATCCTCACC 24760

Qy    301 tttaacactgttgtagatgctcatttt 328
          |||| |||||||| ||| |||
Db  24761 CCGCACACTGGGATAGATGCTTGTTTGT 24788
```

```
RESULT 6
AC026620/c
LOCUS       AC026620   186809 bp   DNA          HTG     18-NOV-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-443G13 map 17, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC026620
VERSION     AC026620.3  GI:11225463
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 186809)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 17, clone RP11-443G13
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 186809)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 18, 2000 this sequence version replaced gi:7534102.
            All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
              Center project name: L8629
              Center clone name: 443_G_13
            --------------- Summary Statistics
              Sequencing vector: M13; M77815; 42% of reads
              Sequencing vector: Plasmid; n/a; 58% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 182357 bases at least Q40
              Consensus quality: 184169 bases at least Q30
              Consensus quality: 184860 bases at least Q20
              Insert size: 182000; agarose-fp
              Insert size: 185609; sum-of-contigs
              Quality coverage: 10.1 in Q20 bases; agarose-fp
              Quality coverage: 9.9 in Q20.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

```
            * be preserved.
            *        1   3134: contig of 3134 bp in length
            *     3135   3234: gap of   100 bp
            *     3235   6135: contig of 2901 bp in length
            *     6136   6235: gap of   100 bp
            *     6236  10291: contig of 4056 bp in length
            *    10292  10391: gap of   100 bp
            *    10392  14688: contig of 4297 bp in length
            *    14689  14788: gap of   100 bp
            *    14789  24656: contig of 9868 bp in length
            *    24657  24756: gap of   100 bp
            *    24757  78004: contig of 53248 bp in length
            *    78005  78104: gap of   100 bp
            *    78105  89254: contig of 11150 bp in length
            *    89255  89354: gap of   100 bp
            *    89355 101823: contig of 12469 bp in length
            *   101824 101923: gap of   100 bp
            *   101924 117510: contig of 15587 bp in length
            *   117511 117610: gap of   100 bp
            *   117611 135414: contig of 17804 bp in length
            *   135415 135514: gap of   100 bp
            *   135515 151939: contig of 16425 bp in length
            *   151940 152039: gap of   100 bp
            *   152040 177533: contig of 25494 bp in length
            *   177534 177633: gap of   100 bp
            *   177634 186809: contig of 9176 bp in length.
FEATURES             Location/Qualifiers
     source          1..186809
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-443G13"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..3134
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    3235..6135
                     /note="assembly_fragment
     misc_feature    6236..10291
                     /note="assembly_fragment"
     misc_feature    10392..14688
                     /note="assembly_fragment"
     misc_feature    14789..24656
                     /note="assembly_fragment"
     misc_feature    24757..78004
                     /note="assembly_fragment"
     misc_feature    78105..89254
                     /note="assembly_fragment"
     misc_feature    89355..101823
                     /note="assembly_fragment"
     misc_feature    101924..117510
                     /note="assembly_fragment"
     misc_feature    117611..135414
                     /note="assembly_fragment"
     misc_feature    135515..151939
                     /note="assembly_fragment"
     misc_feature    152040..177533
                     /note="assembly_fragment"
     misc_feature    177634..186809
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    48912 a  41810 c  42730 g  52153 t   1204 others
ORIGIN

Query Match            57.1%;  Score 195.2;  DB 70;  Length 186809;
Best Local Similarity  78.7%;  Pred. No. 2.5e-52;
Matches 258;  Conservative  0;  Mismatches 68;  Indels  2;  Gaps  2;

Qy    2 ttccaggagattacagctccagccacaatgattgaggtggtttgcaacgaccgtctagga 61
```

```
Db  88227  TGCAAGGCAATTCGAGCTCCAGCTAGGATGATCGAGGTTGTTTGCAACGACCGTCTGGGG  88168
           ||  ||  |||  |||  |||||||  |||||  ||||| |||||  ||||||||||||| ||
Qy      62  aagaaagtccgcgttaagtgcaacaccgatgacacacggggacttgaagaaactgata  121
           ||  || ||||||  |||||||||| |||| |||  |||||||  ||||||| ||||||
Db  88167  AAAAAGGTCCACGTTAAATGCAACACGGATGATACCATCGGGGACCTTAAGAAGCTGATT  88108
Qy     122  gcggccaaactggcactgttggaataagatcgttcttaaaaagtggtacacgatttt  181
           ||  ||||||||| ||||||||  |||||  |||| || ||||  |||||||| ||||||
Db  88107  GCAGCCTAAACTGGCACCCGTTGAACACAGATTGTTCCTGAAGAAGTGGTACACAATTTT  88048
Qy     182  aaggaccatgtatctctgggagattatgaaatcccacgatgggatgaaacctggagctttat  241
           |||||||||  ||||||| |||||||||||||  |||||||||||||||||||||||||||||
Db  88047  AAGGACCATGTGTCTGTCTGGGGGACTATGAAATCCACAATGGGATGAACCTGGAGCTTTAT  87988
Qy     242  taccagtagagggaattcctccacttgcccaacctgctttcctctccat-ggctca  300
           |||||  ||||| ||| ||||  ||||||||||| ||  | ||| | |||
Db  87987  TATCAAAGATGAGAA-TCCTCATCTTCCTGCGCCCCACTCTCCTCTCCCATCCTCACC  87929
Qy     301  tttaacactgttgtagatgctcattttt  328
           |||||  ||||||||  |||||||||||||
Db  87928  CCGCACACTGGGATAGATGCTTGTTTGT  87901
```

RESULT 7
AC074146/c

```
LOCUS       AC074146      201602 bp      DNA          HTG         15-JUL-2000
DEFINITION  Mus musculus clone RP23-107G3, WORKING DRAFT SEQUENCE, 38 unordered
            pieces.
ACCESSION   AC074146
VERSION     AC074146.1  GI:9211256
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 201602)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Mouse
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 201602)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (15-JUL-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 0
            Center clone name: RPCI-23_107G3
            -----
            Summary Statistics
            Consensus quality: 134464 bases at least Q40
            Consensus quality: 159739 bases at least Q30
            Consensus quality: 170106 bases at least Q20
            Estimated insert size: 181300; agarose-fp estimation
            Estimated insert size: 197902; sum-of-contigs estimation
            Quality coverage: 3.41 in Q20 bases; agarose-fp estimation
            Quality coverage: 3.12 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 38 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1     1675: contig of 1675 bp in length
            *       1676     1775: gap of unknown bp in length
            *       1776     3291: contig of 1516 bp in length
```

```
*    3292     3391: gap of unknown length
*    3392     4432: contig of 1041 bp in length
*    4433     4532: gap of unknown length
*    4533     6053: contig of 1521 bp in length
*    6054     6153: gap of unknown length
*    6154     7161: contig of 1008 bp in length
*    7162     7261: gap of unknown length
*    7262     8417: contig of 1156 bp in length
*    8418     8517: gap of unknown length
*    8518     9800: contig of 1283 bp in length
*    9801     9900: gap of unknown length
*    9901    10926: contig of 1026 bp in length
*   10927    11026: gap of unknown length
*   11027    12062: contig of 1036 bp in length
*   12063    12162: gap of unknown length
*   12163    14110: contig of 1948 bp in length
*   14111    14210: gap of unknown length
*   14211    15534: contig of 1324 bp in length
*   15535    15634: gap of unknown length
*   15635    17316: contig of 1682 bp in length
*   17317    17416: gap of unknown length
*   17417    20982: contig of 3566 bp in length
*   20983    21082: gap of unknown length
*   21083    24936: contig of 3854 bp in length
*   24937    25036: gap of unknown length
*   25037    28124: contig of 3088 bp in length
*   28125    28224: gap of unknown length
*   28225    30948: contig of 2724 bp in length
*   30949    31048: gap of unknown length
*   31049    34011: contig of 2963 bp in length
*   34012    34111: gap of unknown length
*   34112    37382: contig of 3271 bp in length
*   37383    37482: gap of unknown length
*   37483    41136: contig of 3654 bp in length
*   41137    41236: gap of unknown length
*   41237    45754: contig of 4518 bp in length
*   45755    45854: gap of unknown length
*   45855    48474: contig of 2620 bp in length
*   48475    48574: gap of unknown length
*   48575    52059: contig of 3485 bp in length
*   52060    52159: gap of unknown length
*   52160    55993: contig of 3834 bp in length
*   55994    56093: gap of unknown length
*   56094    59136: contig of 3043 bp in length
*   59137    59236: gap of unknown length
*   59237    63015: contig of 3779 bp in length
*   63016    63115: gap of unknown length
*   63116    66917: contig of 3802 bp in length
*   66918    67017: gap of unknown length
*   67018    72225: contig of 5208 bp in length
*   72226    72325: gap of unknown length
*   72326    80793: contig of 8468 bp in length
*   80794    80893: gap of unknown length
*   80894    88129: contig of 7236 bp in length
*   88130    88229: gap of unknown length
*   88230    95766: contig of 7537 bp in length
*   95767    95866: gap of unknown length
*   95867   104692: contig of 8826 bp in length
*  104693   104792: gap of unknown length
*  104793   113453: contig of 8661 bp in length
*  113454   113553: gap of unknown length
*  113554   121609: contig of 8056 bp in length
*  121610   121709: gap of unknown length
*  121710   132117: contig of 10408 bp in length
*  132118   132217: gap of unknown length
*  132218   143836: contig of 11619 bp in length
*  143837   143936: gap of unknown length
*  143937   153823: contig of 9887 bp in length
*  153824   153923: gap of unknown length
*  153924   174899: contig of 20976 bp in length
*  174900   174999: gap of unknown length
*  175000   201602: contig of 26603 bp in length.
```

FEATURES   Location/Qualifiers

```
source          1. .201602
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="RP23-107G3"
                /clone_lib="RPCI mouse BAC library 23"
BASE COUNT   52434 a  47973 c  47287 g  49607 t  4301 others
ORIGIN

Query Match          36.8%;    Score 125.8;  DB 75;  Length 201602;
Best Local Similarity 59.0%;   Pred. No. 1.3e-29;
Matches 220;  Conservative  0;  Mismatches  62;  Indels  37;  Gaps  2;

Qy       9 agattacagtccagccacaatgattgaggtggttgcaacgaccgtctaggaagaaag 68
           |||||  |||  ||| |||||||||||||||  |  | ||||  || ||||| ||| |||
Db   40330 AGATTTAAGCTGCCACCACAATGATTGAGGCTGCTAACAACAACCTTCTTGGAAAGGAG 40271

Qy      69 tccgcgttaagtgcaaacaccgatgacaccatcgggacttgaagaaactgatagcggccc 128
           || |  ||| |||||| ||||||||||||||| || | ||||||||||| |||||||||||
Db   40270 TCAGGGTTAAATGCAACACAGATGACACCATCAGGAAACTTGAAGAACTTGATAGCGGCCC 40211

Qy     129 aaactggcactgcttggaataagatcgttcttaaaaagtggtacacgatttttaaggacc 188
           ||                        --------------TGGTACAGGAGTTTGAAGGACC 40187
Db   40210 AA---------------------------

Qy     189 atgtatctctggagattatgaaatccacgatgggatgaacctggagcttattaccagt 248
           |  || |||||||||||||| |||||| ||||| ||||| || || ||||||||||||
Db   40186 ACGTGTCTCTGGAGATTATCAAATCCACAGTGGCATGAACC-AGAGCTCTGTGACCAAG 40128

Qy     249 agagggaattcctccacctgccaacttgcttcctctcccatggctcatttaacac 308
           ||||  ||||||||| ||| ||| ||||||| || ||||||||||||||  ||  |  |
Db   40127 AAAGGGAGATTCCTCCTCCTCCCCCAGCCCGCCTTCCTCCTTCCACGTTCATTTGACACT 40068

Qy     309 tgttgtagatgctcattt 327
           || ||| |||||| |||||
Db   40067 GGTAATAGATTCTCATTTT 40049
```

```
RESULT    8
AC069459/c
LOCUS       AC069459   219217 bp   DNA      HTG    02-FEB-2001
DEFINITION  Mus musculus chromosome 11 clone RP23-168P5, WORKING DRAFT
            SEQUENCE, 20 unordered pieces.
ACCESSION   AC069459
VERSION     AC069459.22  GI:12621364
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 219217)
  AUTHORS   Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
            Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
            Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
            Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
            Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
            Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
            Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
            Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
            Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
            Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
            Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
            Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
            Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
            Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
            Worley,K. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 219217)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
```

```
COMMENT     Baylor Plaza, Houston, TX 77030, USA
            On Jan 31, 2001 this sequence version replaced gi:12408440.
            ---------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: MAFO
                Center clone name: RP23-168P5
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 47% of reads
                Chemistry: Dye-terminator Big Dye: 53% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 209384 bases at least Q40
                Consensus quality: 216147 bases at least Q30
                Consensus quality: 219990 bases at least Q20
                Estimated insert size: 209768; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 6.6x in Q20 bases; sum-of-contigs estimation
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1:  56521:   contig of 56521 bp in length
            *    56522:  56621:   gap of unknown length
            *    56622: 104917:   contig of 48296 bp in length
            *   104918: 105017:   gap of unknown length
            *   105018: 149343:   contig of 44326 bp in length
            *   149344: 149443:   gap of unknown length
            *   149444: 167650:   contig of 18207 bp in length
            *   167651: 167750:   gap of unknown length
            *   167751: 183380:   contig of 15630 bp in length
            *   183381: 183480:   gap of unknown length
            *   183481: 190011:   contig of 6531 bp in length
            *   190012: 190111:   gap of unknown length
            *   190112: 196517:   contig of 6406 bp in length
            *   196518: 196617:   gap of unknown length
            *   196618: 201502:   contig of 4885 bp in length
            *   201503: 201602:   gap of unknown length
            *   201603: 204374:   contig of 2772 bp in length
            *   204375: 204474:   gap of unknown length
            *   204475: 206118:   contig of 1644 bp in length
            *   206119: 206218:   gap of unknown length
            *   206219: 207743:   contig of 1525 bp in length
            *   207744: 207843:   gap of unknown length
            *   207844: 208899:   contig of 1056 bp in length
            *   208900: 208999:   gap of unknown length
            *   209000: 210018:   contig of 1019 bp in length
            *   210019: 210118:   gap of unknown length
            *   210119: 211480:   contig of 1362 bp in length
            *   211481: 211580:   gap of unknown length
            *   211581: 212666:   contig of 1086 bp in length
            *   212667: 212766:   gap of unknown length
            *   212767: 214144:   contig of 1378 bp in length
            *   214145: 214244:   gap of unknown length
            *   214245: 215332:   contig of 1088 bp in length
            *   215333: 215432:   gap of unknown length
            *   215433: 216492:   contig of 1060 bp in length
            *   216493: 216592:   gap of unknown length
            *   216593: 218023:   contig of 1431 bp in length
            *   218024: 218123:   gap of unknown length
            *   218124: 219217:   contig of 1094 bp in length.

FEATURES             Location/Qualifiers
     source          1. .219217
```

```
RESULT 9
AC026142/c
LOCUS       AC026142   156342 bp   DNA        HTG        01-JUN-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-601G18 map 3, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC026142
VERSION     AC026142.2  GI:8139197
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156342)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-601G18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156342)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Illev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,C., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,K., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,M.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 1, 2000 this sequence version replaced gi:7264212.
            All repeats were identified using RepeatMasker.
            Smit, A.F.A & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----------- Project Information
            Center project name: L8318
            Center clone name: 601_G_18
            ----------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 148258 bases at least Q40
            Consensus quality: 152213 bases at least Q30
            Consensus quality: 153912 bases at least Q20
            Insert size: 163000; agarose-fp
            Insert size: 155142; sum-of-contigs
            Quality coverage: 4.5 in Q20 bases; agarose-fp
            Quality coverage: 4.7 in Q20 bases; sum-of-contigs
            -----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *     1    3818: contig of 3818 bp in length
            *  3819  3918: gap of    100 bp
            *  3919  8368: contig of 4450 bp in length
            *  8369  8468: gap of    100 bp
            *  8469 14466: contig of 5998 bp in length
            * 14467 14566: gap of    100 bp
            * 14567 19098: contig of 4532 bp in length
            * 19099 19198: gap of    100 bp
            * 19199 25900: contig of 6702 bp in length
            * 25901 26000: gap of    100 bp
            * 26001 34378: contig of 8378 bp in length
            * 34379 34478: gap of    100 bp
            * 34479 44257: contig of 9779 bp in length
            * 44258 44357: gap of    100 bp
            * 44358 54869: contig of 10512 bp in length
            * 54870 54969: gap of    100 bp
            * 54970 64656: contig of 9687 bp in length
            * 64657 64756: gap of    100 bp
            * 64757 75758: contig of 11002 bp in length
            * 75759 75858: gap of    100 bp
            * 75859 98633: contig of 22775 bp in length
            * 98634 98733: gap of    100 bp
            * 98734 126432: contig of 27699 bp in length
            * 126433 126532: gap of    100 bp
            * 126533 156342: contig of 29810 bp in length.
FEATURES             Location/Qualifiers
     source          1..156342
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-601G18"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..3818
                     /note="assembly_fragment"
     misc_feature    3919..8368
```

```
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="RP23-168P5"

BASE COUNT     56664 a   53067 c  51159 g  56401 t   1926 others
ORIGIN

Query Match         36.8%;   Score 125.8;  DB 74;   Length 219217;
Best Local Similarity  69.0%;  Pred. No. 1.3e-29;
Matches  220;  Conservative  0;  Mismatches  62;   Indels  37;  Gaps  2;

Qy       9  agattacagctccagcccacaatgattgagtggttgcaacgaccgtctaggaagaaag   68
            ||-|||||||||| |||||||||| || ||||||| || |||| ||| ||||||||||
Db  194178  AGATTAAGCTGCCACCACCATGATTGAGGCTGCTAACACAACCTTCTTGGAAAGGAAG  194119

Qy      69  tccggcgttaagtgcaacaccgatgacaccatcggggacttgaagaaacgtgatagcggccc  128
            || |-|||| |||||||||||| || |||||||||||||||||||||| |||||||||||||
Db  194118  TCAGGGTTAAATGCAACAACAGATGCAGCACCATCAGAAACTTGAAGAACTTGATAGCGGCCC  194059

Qy     129  aaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatttttaaggacc  188
            |||||||||| |||||||||||||| |||||||| ||||||||||||
Db          --------------------------------------TGGTACAGGAGTTTGAAGGACC  194035

Qy     189  atgtatcctggagagattatgaaatcccacgatggatgaacctgagctttattaccagt  248
            ||||||||||||||||||| |||||| |||||||||||||| |||||||||||| |||
Db  194034  ACGTGTCTCTGGGAGATTATCAAATCCACAGTGGGATGAACC-AGAGCTCGTGACCAAG  193976

Qy     249  agagggaattcctccaacctggcccaaccttgctttcctctccatggctcattaacac   308
            ||||| |||||||||||||| ||  |||| ||||||||||| |||||||| |||||
Db  193975  AAAGGGAGATTCCTCCTCCTCCCCCAGCCCGCCTCCTTCCTTCCACGTTCATTTGACACT  193916

Qy     309  tgtgtagatgctcatttt  327
            ||-||||| |||||||||
Db  193915  GGTAATAGATTCTCATTTT  193897
```

```
        misc_feature     8469..14466
                         /note="assembly_fragment"
        misc_feature     14567..19098
                         /note="assembly_fragment
                         clone_end:T7
                         vector_side:left"
        misc_feature     19199..25900
                         /note="assembly_fragment"
        misc_feature     26001..34378
                         /note="assembly_fragment"
        misc_feature     34479..44257
                         /note="assembly_fragment"
        misc_feature     44358..54869
                         /note="assembly_fragment"
        misc_feature     54970..64656
                         /note="assembly_fragment
                         clone_end:SP6
                         vector_side:right"
        misc_feature     64757..75758
                         /note="assembly_fragment"
        misc_feature     75859..98633
                         /note="assembly_fragment"
        misc_feature     98734..126432
                         /note="assembly_fragment"
        misc_feature     126533..156342
                         /note="assembly_fragment"
BASE COUNT     47675 a  29500 c  30329 g  47633 t  1205 others
ORIGIN
```

```
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
          Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
          Boguslavkly,L., Boukhgalter,B., Brown,A., Burkett,G.,
          Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
          Collymore,A., Cooke,P., DeArellano,K., Dewar,J.S.,
          Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
          Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
          Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
          Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
          Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
          Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
          McCarthy,M., McEwan,P., McKernan,K., McPheeters,R.,
          Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
          Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
          O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
          Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
          Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
          Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
          Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
          Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
          Young,G., Zainoun,J., Zimmer,A. and.Zody,M.
TITLE     Direct Submission
JOURNAL   Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Jun 7, 2000 this sequence version replaced gi:7677743.
          All repeats were identified using RepeatMasker:
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          Smit, A.F.A. & Green, P. (1996-1997)
          --------------------------------------------------
          Center: Whitehead Institute/ MIT Center for Genome Research
          Center code: WIBR
          Web site: http://www-seq.wi.mit.edu
          Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
          Center project name: LI0205
          Center clone name: 284_K_18
          -------------- Summary Statistics
          Sequencing vector: M13; M77815; 100% of reads
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Assembly program: Phrap; version 0.960731
          Consensus quality: 149867 bases at least Q40
          Consensus quality: 155113 bases at least Q30
          Consensus quality: 157154 bases at least Q20
          Insert size: 163000; agarose-fp
          Insert size: 158526; sum-of-contigs
          Quality coverage: 4.1 in Q20 bases; agarose-fp
          Quality coverage: 4.2 in Q20 bases; sum-of-contigs
          --------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 20 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          --------------
          *          1   1063: contig of 1063 bp in length
          *       1064   1163: gap of    100 bp
          *       1164   3028: contig of 1865 bp in length
          *       3029   3128: gap of    100 bp
          *       3129   5167: contig of 2039 bp in length
          *       5168   5267: gap of    100 bp
          *       5268   8327: contig of 3060 bp in length
          *       8328   8427: gap of    100 bp
          *       8428  11683: contig of 3256 bp in length
          *      11684  11783: gap of    100 bp
          *      11784  13964: contig of 2181 bp in length
          *      13965  14064: gap of    100 bp
          *      14065  18527: contig of 4463 bp in length
          *      18528  18627: gap of    100 bp
          *      18628  22883: contig of 4256 bp in length
          *      22884  22983: gap of    100 bp
          *      22984  29442: contig of 6459 bp in length
          *
```

```
Query Match             35.5%;  Score 121.4;  DB 70;  Length 156342;
Best Local Similarity   71.4%;  Pred. No. 3.4e-28;
Matches  192;  Conservative    0;  Mismatches  66;  Indels  11;  Gaps   2;

Qy       2 ttccaggagattacagctccagccacaatgattgaggtggttgcaacgaccgtctagga  61
           ||| || ||||| |||  |||| ||| |||||||||||||||||||| || ||||||
Db  103294 TTCAAGGTGCGATTGAGCTTTAGCTAGGATGATGGAGGTGTTTGCAACAACTGTCTGGGC 103235

Qy      62 aagaaagtccgcgttaagtgcaacaccgatgacacatcgggagactg---------aa 111
            | ||| |||||| ||||||  ||||| ||||| ||| |||| ||        ||
Db  103234 TTCAG-GTCCAtGTTAAATGCAACATTAGATGACGCCACTGGAGACCTTAAGAGACCGTAA 103176

Qy     112 gaaactgatagcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggta 171
            |||||| || |||  |||| ||||| ||| ||||| |||| ||| |||| ||||||||
Db  103175 GAAGCTGATTGCAGCCCAAGCCCAAGCTGTCACCCATTGTAAAAAGATTATCCTGAAGAAATGGTA 103116

Qy     172 cacgatttttaaggaccatgtatctctcggggagatttatgaaatccacgatgggatgaacct 231
           || |||||||| |||||  ||||||||||||||  |||| ||||||||||| ||| |||||
Db  103115 CATGATTTTTTAGGGACCACCACGATACTGGGGGACCCATGAAATCCATGATGGGGATGGACCT 103056

Qy     232 ggagcttttattaccagtagagggaattc 260
           ||||| ||| |||||   |||| ||||
Db  103055 GGAGCTTTATTTCAATAGATTAGAATCC 103027
```

```
RESULT 10
AC068372/c
LOCUS       AC068372   160426 bp   DNA   HTG   07-JUN-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-284K18 map 3, WORKING DRAFT
            SEQUENCE, 20 unordered pieces.
ACCESSION   AC068372
VERSION     AC068372.2  GI:8315072
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 160426)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 3, clone RP11-284K18
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 160426)
```

```
BASE COUNT     48454 a    31275 c    30247 g    48545 t     1905 others
ORIGIN

Query Match          35.5%;    Score 121.4;   DB 74;    Length 160426;
Best Local Similarity   71.4%;   Pred. No. 3.4e-28;
Matches  192;   Conservative   0;   Mismatches  66;   Indels  11;   Gaps   2;

Qy     2  tccaggagattacagctccagccacaatgatgagtggttgcaacgaccgtctagga  61
          ||| ||||||||||| ||| |||||| || ||||||||||||| || |||| ||||||
Db  70075  TTCAAGGTGATTTGAGCTTAGCTAGGATGATGAGGTTGTTTGCAACAACTGTCTGGGC  70016

Qy    62  aagaaagtccggttaagtgcaacaccgatgacaccatcgggacttg--------aa  111
          |||||||||||||||||||||||| |||||||||||||||||||        ||
Db  70015  TTCAG-GTCCAGTTAAATGCAACATAGATGACGCCACTGGAGACCTTAAGAGACCGTAA  69957

Qy   112  gaaactgatagcggccaaactgcactgtgttggaataagatcgttcttaaaagtggta  171
          ||| ||||| ||||||||| ||| |||||| ||| ||| |||| || |||| ||||||
Db  69956  CATGATTTTTAGGGACCACCACGATACTGGGGACCATGAAATCCATGATGGGATGGACCT  69837

Qy   172  cacgattttaaggacacatgtatcctctgggagattatgaaatccacgatgggatgaacct  231
          ||| |||||||||||||||||||| ||||||||||||| |||||||||||||||||||||
Db  69896  CATGATTTTTAGGGACCACCACGATACTGGGGACCATGAAATCCATGATGGGATGGACCT  69837

Qy   232  ggagcttattaccagtagaggggaattc  260
          ||||||||||||| || ||||| |||||
Db  69836  GGAGCTTTATTTTCAATAGATTAGAATCC  69808

RESULT 11
AB023032/c
LOCUS       AB023032    59762 bp    DNA               PLN    27-DEC-2000
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K5J14.
ACCESSION   AB023032 BA000015
VERSION     AB023032.1 GI:4220631
KEYWORDS    .
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui TAC
            clone:K5J14.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (sites)
  AUTHORS   Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H.
            and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 5. X.
            Sequence features of the regions of 3,076,755 bp covered by sixty
            P1 and TAC clones
  JOURNAL   DNA Res. 7 (1), 31-63 (2000)
  MEDLINE   20181125
REFERENCE   2 (bases 1 to 59762)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-1999) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=K5J14
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
```

```
*     29443  29542: gap of      100 bp
*     29543  36305: contig of 6763 bp in length
*     36306  36405: gap of      100 bp
*     36406  43375: contig of 6970 bp in length
*     43376  43475: gap of      100 bp
*     43476  49721: contig of 6246 bp in length
*     49722  49821: gap of      100 bp
*     49822  56627: contig of 6806 bp in length
*     56628  56727: gap of      100 bp
*     56728  66817: contig of 10090 bp in length
*     66818  66917: gap of      100 bp
*     66918  78729: contig of 11812 bp in length
*     78730  78829: gap of      100 bp
*     78830  92103: contig of 13274 bp in length
*     92104  92203: gap of      100 bp
*     92204 104569: contig of 12366 bp in length
*    104570 104669: gap of      100 bp
*    104670 118863: contig of 14194 bp in length
*    118864 118963: gap of      100 bp
*    118964 133390: contig of 14427 bp in length
*    133391 133490: gap of      100 bp
*    133491 160426: contig of 26936 bp in length.

FEATURES            Location/Qualifiers
     source          1..160426
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-284K18"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1063
                     /note="assembly_fragment"
     misc_feature    1164..3028
                     /note="assembly_fragment"
     misc_feature    3129..5167
                     /note="assembly_fragment"
     misc_feature    5268..8327
                     /note="assembly_fragment"
     misc_feature    8428..11683
                     /note="assembly_fragment"
     misc_feature    11784..13964
                     /note="assembly_fragment
                     /clone_end:SP6
     vector_side:right"
     misc_feature    14065..18527
                     /note="assembly_fragment
                     /clone_end:T7
     vector_side:left"
     misc_feature    18628..22883
                     /note="assembly_fragment"
     misc_feature    22984..29442
                     /note="assembly_fragment"
     misc_feature    29543..36305
                     /note="assembly_fragment"
     misc_feature    36406..43375
                     /note="assembly_fragment"
     misc_feature    43476..49721
                     /note="assembly_fragment"
     misc_feature    49822..56627
                     /note="assembly_fragment"
     misc_feature    56728..66817
                     /note="assembly_fragment"
     misc_feature    66918..78729
                     /note="assembly_fragment"
     misc_feature    78830..92103
                     /note="assembly_fragment"
     misc_feature    92204..104569
                     /note="assembly_fragment"
     misc_feature    104670..118863
                     /note="assembly_fragment"
     misc_feature    118964..133390
                     /note="assembly_fragment"
     misc_feature    133491..160426
                     /note="assembly_fragment"
```

```
FEATURES             Location/Qualifiers
     source          1..59762
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="K5J14"
                     /clone_lib="Mitsui TAC"
     CDS             join(6320..6469,6562..6764,6945..7020,7279..7365,
                     7444..7542)
                     /note="contains similarity to tetracycline transporter
                     protein
                     gene_id:K5J14.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10194.1"
                     /db_xref="GI:10177006"
                     /translation="MKVFSADFQADNVPEGRRASAFGILTGITSCAFVCANLSARFLS
                     ICVTFQVAAAMGLSTLYMRLFLPDSIRDNSLGAPIVINEKLSSPLLEDCPGHRNRIF
                     RAIRLVHEMASLMRSSVPLFQVAMVSFFSSLAEAGLHASSMYLKAKFHFNKDQFADL
                     MIIVGSGSISQLFMPVLVPALKEERLLSIGLFFGCAHVRTVF"
     CDS             join(11437..11634,11729..11846,12151..12249,12342,..12457,
                     12552..12638,12814..13008,13096..13165,13238..13392,
                     13524..13688,13766..13837,13909..14026,14104..14346,
                     14420..14558,14652..14843,14965..15495,15720..15857,
                     15937..15999)
                     /note="gene_id:K5J14.2
                     pirl|T30561
                     similar to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10195.1"
                     /db_xref="GI:10177007"
                     /translation="MRSFFCKMNLQALCFLNFNTLGFMEDQPINQCSSSTNASEKTPE
                     STELNIKTLDSRTYTFQVNKNETVLLFKEKIASETGVPVGQQRLIFRGRVLKDDHPL
                     SEYHLENGHTLHLIVRQPAESAPSSGTPSQGATANDGNNTNGGPSRNGRHVSHSVVLG
                     SFNVGDGTEGIYPDLSRVIGAVLNSEGVQGQLPTNHSTNGTQSSMPSNQSSNAPPGNT
                     SDGEPGIGGGQSQATGHSQPRQAFPGVSFQTSMPRVVQIPVTAATTIPSFLTPIPDS
                     LDTLMEFINRMEQALSQNGYQPDTSSAGSGGRPREELPRNRGAATPEALSVVLRNAQ
                     HLLSGLGVSSLSHIAGRLEQDGSSSSDPTLRSQIOTEAVQLAMQHLGALLLELGRTI
                     LTLRMAPSPELSYVNAGPAVYISPSGPNPIMVQPFPHOISPLFTGATVSSNPLTGPVG
                     LGTAQRHININIHAGTSGSPMLSSVGNQRSNEGGQGDRDSNTSSVPAAVPSHSTGEN
                     VSAGVQPGLGGNVSVAQINARIRDMVNIMGRDQIPSGIESLERDMSTGHGVATAMPE
                     QPTNIATTCAPESSGGSGSIHDLPSERSNSVCQNEKDLGGDLEHPARAKDTSCTTGQSSA
                     PSGDATGDAKETNKATPEVAIATPLGLGLGGLDRKKRSKQPKVSGKTEDSGTSATLEG
                     VQQSSGTSGQQLQSLFSGSSRSDETGLRRCQCSDDRVDVSSAMSQVLESPVLDGLLA
                     GVSRQAGVDSPNMLRNMLQQFTONPQIMNTVQQIAQQVDGQEIENMMSGGAQGEGGGF
                     DFSRMVQQMMPLVSRAFSQGGPLPHPATIQPDDRQPSOVNVQSMAQMIEHSDPPEDVF
                     RAMVENAAISQDELVNELCCDEALSQEYAELLRRDIEGRLKDDQGL"
     CDS             join(20293,..20530,20628..20735,20827..20919,21008,..21100,
                     21179..21264,21337..21615,21684..21817,21912..22004,
                     22093..22210,22294..22461)
                     /note="gene_id:K5J14.3"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="serine carboxypeptidase-II like"
                     /protein_id="BAB10196.1"
                     /db_xref="GI:10177008"
                     /translation="MAIVSLRDVAMVMTVQVFARGYPETDLVVRLPCQQPKVVFRQYA
                     GYVDLDLNAGRSLFYYFVEAEKHPDTKPLTLWLNGGPGCSSVGGAFTELGPFEYPTGY
                     GRGLRINSMSWNKASNLLFVDSPAGVWGSYSNRSSDVNAGDKSAASDMLVFLLRWFDK
                     FPELKSHDLFLTGESYAGHYIPQLADAILSYNSRSSGFKFNIKGIAIGNPLLKLDRDI
                     PAVYEFWSHGMISEVVGRTIKIQCDFSHYTYAYPHNVSDACNDAIREAGDITTEYVN
                     TFDVLPDLCYPSIALQELRLKQMATKMSMGVDVCMNYERQFYLNIPEVQMALHANRTN
                     LPYSWSLCSNLLNYSAIDVNTRNMLPTLKRIIQNKIPVRIFSGQDSVVPFLGTRTIVG
                     ELANDLNFKTTVPYGVWFHKRQVGGNAIEYGNLLTFATVRGAAHAVAYTQPSRALHF
                     STFLRGQRLPNKTDIAMHD"
     CDS             join(27261,27510,27591,..27698,27785,..27877,27952,..28044,
```

right column:

```
                     28143..28228,28314,..28592,28670,..28803,28922,..29014,
                     29085,..29202,29308..29475)
                     /note="gene_id:K5J14.4"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="serine carboxypeptidase II-like"
                     /protein_id="BAB10197.1"
                     /db_xref="GI:10177009"
                     /translation="MASVSWRAVAVAMVVLLSLQWFAKGYPEEDLVVRLPGQPTVGF
                     KQYAGYVDVKAGRSLFYYYEAVKQPDSKPLTLWLNGGPGCSSIGGGAFTELGPFY
                     PTGDGRGLRVNSMSWNKASHLLFVESPAGVGWSYSNKSSDYNTGDKSTANDMLVFLLR
                     WFEKFPKLKSRDLFLTGESYAGHYIPQLADAILSYNSHSSGFKFNIKGVAIGNPLLKL
                     DRDSPATYEFWSHGMISDELKLTITSQCFDDYTFASPHNVSTACNEAISETENIIT
                     EYVNNTDVLLDVCYPSIVQQELRLKKMATKMSMGVDVCMTYERRFYFNLPEVQKALHA
                     NRTHLPYSWSMCSGVLNYSDIDGNIDMLPILKRIILNKTPIWIFSGQDSVVPFGGSR
                     TLVRELAQDLNFKTTVPYGAWFHKSQVGGWAIEYGKLLTFATVRGAAHMVPYAQPSRA
                     LHLFSSFVSGRRLPNNTHSSTDE"
     CDS             join(32631,..32691,32788..32924,33757..33803,34429,..34757,
                     34840..34922,35182..35257,35373..35437,35509..35604,
                     35716..35994)
                     /note="gene_id:K5J14.5"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="alcohol dehydrogenase"
                     /protein_id="BAB10198.1"
                     /db_xref="GI:10177010"
                     /translation="MENGNSSSDDNKSSHKPIRCKAAVSRKAGEPLVMEEIMVAPPQPF
                     EVRIRIICTALCHSDVTFWKLQVPPACFPRILGHEAIGVVESVGENVKEVVEGDTVLP
                     TFMPDCGDCVDCKSHKSNLCSKFPFKVSPWMPRYDNSSRFTDLNGETLFHFLNVSSFS
                     EYTVLDVANVVKIDSSIPPSRACLLSCGVSTGVGAAWETAKVEKGSTVVIFGLGSIGL
                     AVAEGARLCGASRIIGVDINPTKFQVGQKFGVTEFVNSMTCEKNRVSEVINEMTDGGA
                     DYCFEOCVGSSSLVQEAYACCROGWGKTITLGVDKPGSQICLDSFDVLHHGKILMGSLF
                     GGLKAKTHIPLLKRYLSNELELDKFVTHEMKFEEINDAFQLLLEGKCIRCVLWMG"
     CDS             37256..38779
                     /note="gene_id:K5J14.6"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="beta-glucosidase"
                     /protein_id="BAB10199.1"
                     /db_xref="GI:10177011"
                     /translation="MRTIYLSLLVFIIVLALNEVMAKKHSSTPKLRRSDFPEDFIFGA
                     ATSAYQEGAAHEDGRGPSIWDTFSEKYPEKIKDGSNGIASDSYHLYKEDVGLLHQI
                     GFDAYRFSISWSRILPENLKGGINQAGIDYYNNLINELLSKGIKPFATIFHWDTPQS
                     LEDAYGGFLGAEIVNDFRDYADICFKNFGDRVKHWMTLNEPLTVVQOGYVAGVMAPGR
                     CSKFTNPNCTAGNSATEPYIVGHNLLLAHGEAVKVYREKYKASQKGQVGIALNAGWNL
                     PYSESAEDRLAAARAMAFTFDYFMEPLVTGKYPIDMVNVKGGRLPTTAKQSKMLKG
                     SYDFIGRNYYSSYAKDVPCSSENVTLFSDPCASVTDGEREGVPIGPRKAASDWLLIYPK
                     GIRDLLYAYSSYAKDVPITENGRDEASTGHGVATAMPE
                     ANVKGFFAWSLLDNFEWATGYAVRFGLYYVDFNGGRKRYPKKSAKWFKLLNEKKKN"
     CDS             join(41203..42492,42572,..42818,42900..43083,43173,..43344,
                     43425,..43646)
                     /note="gene_id:K5J14.7"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="cell division protein FtsH"
                     /protein_id="BAB10200.1"
                     /db_xref="GI:10177012"
                     /translation="MATTSSNPLLLSSNFLGSGOIIISAPTPKTTTKSLPFSVISRKRY
                     QISQSEKLMKSLPSQAALAALLFSSSSPQALAVNEPVQPPAPTITAEAQSPNLSTFGQ
                     NVLMTAPNPQAQSSDLPDGTQWRYSEFLNAVKKGKVERVKVFSKDGGSVLQLTAVDNRRA
                     TVIVPNDPDLIDILAMNGVDISVSEGEGNGLDFFIGNLLFLLPLAFGGLFYLPRGGQG
                     GAGGPGGLGGPMDFGRSKSKFQEVPETGVTFGDVAGADQAKLELQEVVDFLKNPDKYT
                     ALGAKIPKGCLLVGPPGTGKTLLARAVAGEAGVPFFSCCAASEFVELFVGVGASRVRDL
                     FEKAKSKAPCIVFIDEIDAVGRQRGAGMGGGNDEREQTINQLLTEMDGFSGNSGVIVL
                     AATNRPDVLDSALLRPGRFDRQVTVRDPDVAGRVQOLKVHSRGKAIGKDVDYEKVARR
                     TPGFTGADLQNLMNEAAILAARRELKEISKDEISDALERIIAGPEKKNAVVSEKKRL
                     VAYHEAGHALVGALMPEYDPVAKISIIPRGQAGGLTFAPSEERLESCLYSRSYELNQ
                     MAVALGGRVAEEVIFGDENVTGASNDFMQVSRVARQMVERFGFSKKIGQVAVGGAGG
                     NPFLGQSMSSQKDYSMATADVDAEVRELVEKAYVRAKEIITTQIDILHKLAQLLIEK
                     ETVDGEEFMSLRFIDGQAELYVS"
                     complement(join(44029..45531,45787..45882,46066..46551))
                     /note="contains similarity to CHP-rich zinc finger
                     protein
```

(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MJC20 and the 3' clone is MDH9.

```
                 gene_id:K5J14.8"
                 /codon_start=1
                 /evidence=not_experimental
                 /protein_id="BAB10201.1"
                 /db_xref="GI:1017013"
                 /translation="MEAPSPFFSRSKPKCRLSDPAQPLNLFRLGCSSSKGNDMEDTPP
                 PFFSTSKPKCRLSDPAQPHKLCRRGYRSHPRLSMCFSCKGKDMEDSRYYYCVTCKLE
                 FHRGCHVFPPEIKHIFHPSHPLTLISVDPTFDFSPRNWKDKSSTEESIYEVDDDCEDD
                 DNDDDDDDDDDNDDDVVDDDDDDDDDDVVDDDYGDSACLSDGNHLKCKCCQVP
                 LQRVYYHCSICKFLNATCSMRPPPTISHIKSHEHTLTLFPIRLPFPCDACGLSLSE
                 TDDLVYACLPCSHHVHRSCIYLPRVIKITRHPHRLSFSLSLQPGDFLCGVCRQTIDVN
                 NGQYSCNKGCHYAVHSKCATRKEVWDGKDLEGVPEEPEEYMESFVRIDEETIQHFSHE
                 HYLKLYEKNNICEKDKVCEACTLPVMISQRYYGCMKCDFVLDEACAFLPRKKYHPLHK"

Query Match     34.7%;   Score 118.8;  DB 12;  Length 59762;
Best Local Similarity   68.2%;   Pred. No. 2e-27;
Matches  165;  Conservative  77;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     29 atgattgagtggtttgcaacgaccgtctaggaagaaagatccggttaagtgcaacacc 88
           ||||||||||||| || |||||||||||||| ||||||| |||||| || ||||| ||
Db  52376 ATGATCCGAGGTGGTTCTCAACGATCGTTTAGGGAAAAAAGTTAGGTGAACGTGTAACGAT 52317

Qy     89 gatgcacccaccgtcggggacttgaagagaaactgatagcggcccaaactggcactcgttggaat 148
           ||||||| |||| |||| |||| || ||||| ||||  |||| ||||| |||| || || || ||
Db  52316 GATGACACCATCGGTGATCTGAAGAAGCTTGTCCGCGGCACAAACCGGACGAGCCGAG 52257

Qy    149 aagtcgttcttaaaagtggtacacgatttttaaggaccatgtatctctcggagattat 208
           |||| | || ||  |  |||||| |||| || ||| ||| ||||| | || ||  ||||
Db  52256 AAGATCAGAATTCAGAAGTGGTACAACATCTACAAGGATCACATCCACTCTCCAAGGACTAT 52197

Qy    209 gaaatcccacgatgggatgaacctgagcttattaccagtagagggaattcctccacct 268
           ||||  ||| || || ||||| |||| |||  ||| ||  | || ||  || ||| |||
Db  52196 GAGATCCATGACGGCATGGGTGCTTCTGAGCTTTACAACTAGGTTTTACATCTTTCTTAT 52137

Qy    269 tg 270
           ||
Db  52136 TG 52135

RESULT 12
ATT14D3/c
LOCUS       ATT14D3       88010 bp    DNA            PLN       04-FEB-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 3, BAC clone T14D3.
ACCESSION   AL138649
VERSION     AL138649.1  GI:6899911.
KEYWORDS    .
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 88010)
AUTHORS     Jordan,N., Bangert,S., Wiedelmann,R., Voss,H., Unseld,M.,
            Mewes,H.W., Lemcke,K., Mayer,K.F.X., Quetier,F. and Salanoubat,M.
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 88010)
AUTHORS     EU Arabidopsis sequencing project.
TITLE       Direct Submission
JOURNAL     Submitted (04-FEB-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
            Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
            d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
            Gaston Cremieux, BP191, 91006 Evry Cedex, France;
            http://www.genoscope.cns.fr
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1..88010
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
```

```
     exon            /db_xref="taxon:3702"
                     /chromosome="3"
                     complement(1145..2116)
                     /number=1

     CDS             complement(1145..2116)
                     /note="strong similarity to FLAVONOL 4'-SULFOTRANSFERASE -
                     Flaveria chloraefolia, EMBL:M84136"
                     /codon_start=1
                     /product="sulfotransferase-like protein"
                     /protein_id="CAB72145.1"
                     /db_xref="GI:6911845"
                     /translation="MEMNLRIEDLNEETKTLISSLPSDKDFTGKTICKYQGCWYTHNV
                     LQAVLNFQKSFKPQDTDIIVASPKCGTWLKALTFALLHRSKQPSHDDDHPLLSNNP
                     HVLVPYFEIDLYLRSENPDLTKFESSSPRLFSTHVPSHTLQEGLKGSTCKIVYISRNVK
                     DTLVSYWHFFTKKQTDEKIISSFEDTFEMFCRGVSIFGPFWDHVLSYWRGSLEDPNHV
                     LFMKFEEMKAEPRDQIKKFAEFLGCPFTKEEESSGSVDEIIDLCSLRNLSSLEINKTG
                     KLNSGRENKMFFRKGEVGDWKNYLTPEMENKIDMIIQEKLQNSGLKF"

     gene            1145..2116
                     /gene="T14D3.10"

     gene            3483..4472
                     /gene="T14D3.20"

     CDS             complement(3483..4472)
                     /note="strong similarity to FLAVONOL 4'-SULFOTRANSFERASE -
                     Flaveria chloraefolia, EMBL:M84136"
                     /codon_start=1
                     /product="sulfotransferase-like protein"
                     /protein_id="CAB72146.1"
                     /db_xref="GI:6911846"
                     /translation="MDEKKITMNVRNDELSEESKTLISSLPSDKNSTGVNVCKYQGCW
                     YTPPILQGVLNFQNFKPQDTDIVASFPKCGTWLKALTFALVRRSKHPSHDDHPL
                     LSDNPHVLSPSLEMYLYLCSENPDLTKFSSSSRLFSTHMPSHTLQEGLKGSTCKIVYM
                     SRNVKDTLVSYWHFFCKKQTDDNIISSVEDTFEMFCRGVNFFGPFWDHVLSYWRGSLE
                     DPNIVLFMKFEEMKEEPRQIKRLAEFLGCLFTKEEEESGLVDEIIDLCSLRNLSSLE
                     INKTGKLHSTGRENKTFFRKGEVGDWKNYLTPEMENKIDMIIQEKLQNSGLKF"

     exon            complement(3483..4472)
                     /number=1

     exon            complement(5519..5902)
                     /number=1

     gene            5519..9216
                     /gene="T14D3.30"

     CDS             complement(join(5519..5902,6007..6129,6217..6936,
                     7034..7426,7615..7696,8280..8578,9067..9216))
                     /note="similarity to 2-phosphoglycerate kinase -
                     Methanococcus jannaschii, PIR:A64485; Contains
                     ATP/GTP-binding site motif A (P-loop)
                     AA164-171;Prokaryotic membrane lipoprotein lipid
                     attachment site AA158-168;Prokaryotic membrane lipoprotein
                     lipid attachment site AA386-396"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB72147.1"
                     /db_xref="GI:6911847"
                     /translation="MTEATKVLYIVVREEGDDDNNGDDSFRYTRPVLQSTLQLMGCK
                     ARHAFKISRRVFELIRSEGSCNTSPENGKEPEFAKEVGSTCVEKLNCLVVAGDVDKN
                     KSKPFEMYKRRTTVVVSREIFVDVVCDALAEYKYVGRDQRADLILACRIRERKESVTV
                     LLCGTSGCGKSTLSALLGSRLGITTVVSTDSIRHMMRSFADEKQNPLLWASTYHAGEY
                     LDPVAVAESKAKRKAKLKGSRGVNSNAQKTDAGSNSSTELLSHKQNAIEGYRAQSE
                     MVIDSLDRLITTWEERNESVVVEGVHLSLNFVMGLMKKHSIVPFMVYIANEEKHLER
                     FAVRAKYWTLDPAKNKYKVYIRNIRTIQDYLCKRADKHLVPKINTNVDKSVATIHAT
                     VFGCLRRETGEKLYDTTNTVSVIDDEHRNQCAANSITSKGMFQVIQRQGSRRFMA
                     LCNTDGTVAKTWPVASVGKIRKPVVNTEMDDGTEHQLHKAEPVNLQFGHFGISAWPSD
                     GATSHAGSVDDLRADIIETGSRHYSSCCCSSPRTSDGPSKELMEEQSVNGSDEDDEGD
                     DDFHEPDSDEDLSDNNDERNRDEIGSVNDEDSTKSDEEYDDLAMEDKSYWTDNEEEESR
                     DTISMVSQNNHNEASKTNKDDKYSQNLDLFLKTTNQPLTESLELTSEYRNRMGVAASD
                     KAKMRKRSLSIPPVGKHGSIIDDQILANQTDSVL"

     intron          complement(5903..6006)
                     /number=1

     exon            complement(6007..6129)
                     /number=2

     intron          complement(6130..6216)
                     /number=2

     exon            complement(6217..6936)
                     /number=3
```

```
intron          complement(6937..7033)
                /number=3
exon            complement(7034..7426)
                /number=4
intron          complement(7427..7614)
                /number=4
exon            complement(7615..7696)
                /number=5
intron          complement(7697..8279)
                /number=5
exon            complement(8280..8578)
                /number=5
intron          complement(8579..9066)
                /number=6
exon            complement(9067..9216)
                /number=7
LTR             14402..14473
                /note="72 bp long terminal repeat"
misc_feature    14495..18651
                /note="similarity to fragments of copia-like polyprotein"
LTR             19002..19073
                /note="72 bp long terminal repeat"
exon            19415..19534
                /gene="T14D3.40"
                /number=1
gene            19415..21625
                /gene="T14D3.40"
CDS             join(19415..19534,19838..20410,20508..20662,20758..20837,
                20930..21066,21169..21272,21360..21450,21533..21625)
                /gene="T14D3.40"
                /note="strong similarity to
                n-acetylglucosaminyl-phosphatidylinositol -
                Schizosaccharomyces pombe, PIR:T40367"
                /codon_start=1
                /product="n-acetylglucosaminyl-phosphatidylinositol -like
                protein"
                /protein_id="CAB72148.1"
                /db_xref="GI:6911848"
                /translation="MAEPPKLRVLMVSDFFFPNFGGVENHIYYLSQCLLKLGHKVVVM
                THAYGNRSGVRYMTGGLKVYYVPWRPFVMQTTFPTVIGTLPIVRTILRREKITVVHGH
                QAFSTLCHEALMHARTMGYKVVFTDHSLYGFADVGSIHMNKVLQFSLADIDQAICVSH
                TSKENTVLRSGLSPAKVFMIPNAVDTAMFKPASVRPSTDIITIVVISRLVYRKGADLL
                VEVIPEVCRLYPNVRFVVGGDGPKHVRLEEMREKHSLQDRVEMLGAVPHSRVSVLVT
                GHIFLNSSLTEAFCIATLEAASCGLLTVSTRVGGFLHGLQVLPDDMVVLAEPDPDDMV
                RAIEKAISILPFINPEMHNRMKKLYSWQDVAKRTEIVYDRALKCSNRSLLERLMRFL
                SCGAWAGKLFCMVMILDYLLWRLLQLLQPDEDIEEAPDICLCHHRGVEVSEGLRKKIK
                "
intron          19535..19837
                /gene="T14D3.40"
                /number=1
exon            19838..20410
                /gene="T14D3.40"
                /number=2
intron          20411..20507
                /gene="T14D3.40"
                /number=2
exon            20508..20662
                /gene="T14D3.40"
                /number=3
intron          20663..20757
                /gene="T14D3.40"
                /number=3
exon            20758..20837
                /gene="T14D3.40"
                /number=4
intron          20838..20929
                /gene="T14D3.40"
                /number=4
exon            20930..21066
                /gene="T14D3.40"
                /number=5
intron          21067..21168
                /gene="T14D3.40"
```

```
exon            21169..21272
                /gene="T14D3.40"
                /number=5
                /number=6
intron          21273..21359
                /gene="T14D3.40"
                /number=6
exon            21360..21450
                /gene="T14D3.40"
                /number=7
intron          21451..21532
                /gene="T14D3.40"
                /number=7
exon            21533..21625
                /gene="T14D3.40"
                /number=8
gene            22937..23299
                /gene="T14D3.50"
                /number=1
CDS             22937..23299
                /gene="T14D3.50"


Query Match         32.8%;     Score 111;    DB 13;    Length 88010;
Best Local Similarity 68.6%;    Pred. No. 7.7e-25;
Matches 153;  Conservative 0;  Mismatches 70;  Indels 0;  Gaps 0;

Qy      29 atgattgaggtggttgcaacgaccgctctaggaaagaaagtccgcgttaagtcaacacc 88
           ||||| ||||||| || ||| || ||||||||| || |||| | |||||||||||||||
Db   53979 ATGATAGAGGTGGTGCTTAATGATCGGCTTGGGAAGAAGAGGTGAGAGTGAAGTCAACGAG 53920

Qy      89 gatgacaccatcgggagacttgaagaaactgatagcggcccaaactggcactcgttggaat 148
            ||||||||||||||| ||||||||| |||||||  ||| |||||||||||| ||||||
Db   53919 GAAGACACCATCGGGAGACTCCAAGAAGCTTGTCGGCGCCAAACTGGAACCCGACCCCGAA 53860

Qy     149 aagatcgttcttaaaagtggtacacgatttttaaggaccatgtatcctcgggagattat 208
             || |||||| ||  ||||| ||||| | | |||||||||| || || ||||| |||
Db   53859 AAGATTCGGATCCAAAAAATGGTACAACATCTACAAGGATCATATCCCTCCAAGGACTAC 53800

Qy     209 gaaatccacgatgggatgaacctggagcttattaccagtaga 251
            |||||| || || |||||||| |  |||||| ||| || ||
Db   53799 GAAATCCATGACGACGGCCATGGGTCTCGAGCTCTACAACTAAA 53757


RESULT 13
AC016808/c
LOCUS       AC016808               182320 bp    DNA             HTG       22-JUL-2000
DEFINITION  Homo sapiens clone RP11-11G12, WORKING DRAFT SEQUENCE, 17 unordered
            pieces.
ACCESSION   AC016808
VERSION     AC016808.4  GI:9369471
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182320)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-11G12
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 182320)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkly,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
```

```
TITLE      Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
           Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
JOURNAL    Direct Submission
           Submitted (07-DEC-1999) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Jul 22, 2000 this sequence version replaced gi:8137597.
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           -----------Genome Center
           Center: Whitehead Institute/ MIT Center for Genome Research
           Center code: WIBR
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence_submissions@genome.wi.mit.edu
           ------------ Project Information
           Center project name: L3153
           Center clone name: 11_G_12
           ------------ Summary Statistics
           Sequencing vector: M13; M77815; 95% of reads
           Sequencing vector: Plasmid: n/a; %-0.f%% of reads
           4.620462040462Chemistry: Dye-terminator Big Dye; 100% of
reads
           Assembly program: Phrap; version 0.960731
           Consensus quality: 172181 bases at least Q40
           Consensus quality: 177293 bases at least Q30
           Consensus quality: 179165 bases at least Q20
           Insert size: 185000; agarose-fp
           Insert size: 180720; sum-of-contigs
           Quality coverage: 4.1 in Q20 bases; agarose-fp
           Quality cov.
         * NOTE: This is a 'working draft' sequence. It currently
         * consists of 17 contigs. The true order of the pieces
         * is not known and their order in this sequence record is
         * arbitrary. Gaps between the contigs are represented as
         * runs of N, but the exact sizes of the gaps are unknown.
         * This record will be updated with the finished sequence
         * as soon as it is available and the accession number will
         * be preserved.
         *     1  1052: contig of 1052 bp in length
         *  1053  1152: gap of  100 bp
         *  1153  2627: contig of 1475 bp in length
         *  1153  2627: gap of  100 bp
         *  2628  2727: gap of  100 bp
         *  2728  6363: contig of 3636 bp in length
         *  6364  6463: gap of  100 bp
         *  6464 10575: contig of 4112 bp in length
         * 10576 10675: gap of  100 bp
         * 10676 16048: contig of 5373 bp in length
         * 16049 16148: gap of  100 bp
         * 16149 22232: contig of 6084 bp in length
         * 22233 22332: gap of  100 bp
         * 22333 28640: contig of 6308 bp in length
         * 28641 28740: gap of  100 bp
         * 28741 35908: contig of 7168 bp in length
         * 35909 36008: gap of  100 bp
         * 36009 41361: contig of 5353 bp in length
         * 41362 41461: gap of  100 bp
         * 41462 48668: contig of 7207 bp in length
         * 48669 48768: gap of  100 bp
         * 48769 58942: contig of 10174 bp in length
         * 58943 59042: gap of  100 bp
         * 59043 68408: contig of 9366 bp in length
         * 68409 68508: gap of  100 bp
         * 68509 81914: contig of 13406 bp in length
         * 81915 82014: gap of  100 bp
         * 82015 93085: contig of 11071 bp in length
         * 93086 93185: gap of  100 bp
         * 93186 108502: contig of 15317 bp in length
         * 108503 108602: gap of 100 bp
         * 108603 129972: contig of 21370 bp in length
         * 129973 130072: gap of 100 bp
         * 130073 182320: contig of 52248 bp in length.
FEATURES            Location/Qualifiers
     source          1..182320

     misc_feature    /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-11G12"
                     /clone_lib="RPCI-11 Human Male BAC"
                     1..1052
     misc_feature    /note="assembly_fragment"
                     1153..2627
     misc_feature    /note="assembly_fragment"
                     2728..6363
     misc_feature    /note="assembly_fragment"
                     6464..10575
     misc_feature    /note="assembly_fragment"
                     10676..16048
     misc_feature    /note="assembly_fragment"
                     16149..22232
     misc_feature    /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
                     22333..28640
     misc_feature    /note="assembly_fragment"
                     28741..35908
     misc_feature    /note="assembly_fragment"
                     36009..41361
     misc_feature    /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
                     41462..48668
     misc_feature    /note="assembly_fragment"
                     48769..58942
     misc_feature    /note="assembly_fragment"
                     59043..68408
     misc_feature    /note="assembly_fragment"
                     68509..81914
     misc_feature    /note="assembly_fragment"
                     82015..93085
     misc_feature    /note="assembly_fragment"
                     93186..108502
     misc_feature    /note="assembly_fragment"
                     108603..129972
     misc_feature    /note="assembly_fragment"
                     130073..182320
     misc_feature    /note="assembly_fragment"

BASE COUNT     56034 a   31844 c   31326 g   61515 t   1601 others
ORIGIN

Query Match             32.5%;  Score 111;  DB 65;  Length 182320;
Best Local Similarity   75.3%;  Pred. No. 8.8e-25;
Matches  165;  Conservative   0;  Mismatches  50;  Indels   4;  Gaps   2;

Qy     1 gttccaggagattacagctccagccacaatgattgaggtggtttgcaacgaccgtctagg 60
           ||-||||||||||||||||||||| |||-|||||||||||||| ||||||| || || ||
Db 69304 GTACAAGTGATTCCAGCTCCAGCCACCATGACCAAGATTGTTTGC-ACGACTTTCTGG 69246

Qy    61 aaagaaagtccgcgttaagtgcaacaccgatgacaccatcggggacttgaagaaactgat 120
           |||| ||||| ||||||||||||||||||||| ||||||||||| ||||| ||||| |||
Db 69245 GAAGAAGTCCCGCCATTAAATGCAACACGGATGACAACACCATCGGGGACCCTAAGAAGCTGAT 69186

Qy   121 agcggcccaaactggcactcgttggaataagatcgttcttaaaagtggtacacgatttt 180
           || ||||||| |||| ||| | | |||| |  ||||| -|| ||| ||||||||| |||||
Db 69185 TGCGTCCCAAACTGGCACCCTGCTGGACAAGAGTGTCCTGGAACAA---GTACACTATTTT 69129

Qy   181 taaggaccatgtatctctggagattatgaaatccacga 219
           |||-|| | |||||  ||||||||||| |||||||| | |
Db 69128 TAACAATCACATGTCCTTGGGGACTATGAAATTGAGGA 69090

RESULT 14
AC068782_4
WPCOMMENT
Sequence split into 5 fragments   LOCUS AC068782   Accession AC068782
      Fragment Name        Begin         End
      AC068782_0              1        110000
```

```
AC068782_1    100001    210000
AC068782_2    200001    310000
AC068782_3    300001    410000
AC068782_4    400001    494838
Continuation (5 of 5) of AC068782 from base 400001 (AC068782 Mus musculus chromosome 11
```

```
Query Match            31.9%;    Score 109; DB 74;   Length 94838;
Best Local Similarity  65.8%;    Pred. No. 3.5e-24;
Matches 194;  Conservative  0;  Mismatches 65;  Indels 36;  Gaps  1;

Qy    33 ttgaggtggtttgcaacgaccgctctaggaaagaaagtccgcgttaagtgcaacaccgatg 92
         | |||| ||| ||||  || | |||| |||||| | |||||| ||||||||||| ||||||
Db 68927 TAGAGGCTGCTAACAACAACCTTCTTCTTGGAAAGGAAGTCAGGGTTAAATGCAACACAGATG 68986

Qy    93 acaccatcgggagacttgaagaaactgatagcggcccaaactggcactcgttggaataaga 152
         |||||||  ||| ||||||||||  |||||||||||||||
Db 68987 ACACCATCAGAAAACTTGAAGAACTTGATAGCGGCCCAA------------------- 69024

     153 tcgttcctaaaaagtggtacacgatttttaaggaccatgtatcctggagattatgaaa 212
         ||||||   |||| |||  |||   |||| | |||||||||| | |||||||||| |||
Db 69025 ---------TGGTACAGGAGTTGAAGGACCACCGTGTCTCTGGGAGATTATCAAA 69070

Qy   213 tccacgatggatgaacctggagctttattaccagtagaggggaattcctccacctgcc 272
         ||||| || ||||||||  |||| | |||||||||||| |  |||||||||| || ||
Db 69071 TCCACAGTGAGGATGAACCAGAGCTCTGTGACCAAGAAAGGGAGAATTCCTCCTCCTCCCC 69130

Qy   273 caacctgctttcctctcccatggctcattaacactgttgatgctctcatttt 327
         || || ||| |||| ||| ||| || ||| ||||||| |  || || ||||
Db 69131 CAGCCCTGCTTCCTTCCTTCCACGTTCATTTGACACTGTAATAGATTCATTTT 69185
```

```
Query Match            31.6%;    Score 108; DB 6;    Length 439;
Best Local Similarity  69.3%;    Pred. No. 2.9e-24;
Matches 147;  Conservative  0;  Mismatches 65;  Indels  0;  Gaps  0;

Qy    50 gaccgtctaggaaagaaagtccgcgttaagtgcaacaccgatgacaccatcgggggacttg 109
         || | ||| ||||| |||||| ||| |||||||| |||| | | ||| |||| ||| |||
Db    63 GATAGGCTTGGGAAAAAGGTTCGGGTTAAATGTAACCCGACACAGATAAAGTCGGTGATTTG 122

Qy   110 aagaaactgatagcggcccaaactggcactcgttggaataagatcgttcttaaaagtgg 169
         | ||||||||| || |||||||||| | | | || |||||| | | |||| ||||| ||
Db   123 AAAAAACTGATTGCTGCCCAAACTGGTACCGCTCCGGAACGTATTGTTCTAAAGAAATGG 182

Qy   170 tacacgattttaaggaccatgtatctctggagattatgaaatccacgatgggatgaac 229
         |||||  ||| |||| |||||| | ||||| | |  ||||||| | |  ||||||||||
Db   183 TACACCATTTACAAAGACCATGTCACTCTGCCGCGACTACGAAATCAATGACGGAATGAAC 242

Qy   230 ctggagctttattaccagtagagggggaattcc 261
         ||||| | | ||||||||| |  |  | ||| |
Db   243 TTGGAGTTATATTATCAGTAAAACTGTCTCCC 274
```

Search completed: June 13, 2001, 20:25:44
Job time: 6587 sec

```
RESULT 15
FHU10292
LOCUS       FHU10292        439 bp    mRNA             INV       29-JUL-1999
DEFINITION  Fasciola hepatica clone Fas 3-4 trans-spliced mRNA with spliced
            leader.
ACCESSION   U10292
VERSION     U10292.1  GI:499274
KEYWORDS    .
SOURCE      liver fluke.
ORGANISM    Fasciola hepatica
            Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
            Trematoda; Digenea; Echinostomida; Echinostomata; Fascioloidea;
            Fasciolidae; Fasciola.
REFERENCE   1  (bases 1 to 439)
AUTHORS     Davis,R.E., Singh,H., Botka,C., Hardwick,C., Ashraf el Meanawy,M.
            and Villanueva,J.
TITLE       RNA trans-splicing in Fasciola hepatica. Identification of a
            spliced leader (SL) RNA and SL sequences on mRNAs
JOURNAL     J. Biol. Chem. 269 (31), 20026-20030 (1994)
MEDLINE     94327554
REFERENCE   2  (bases 1 to 439)
AUTHORS     Davis,R.E.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1994) Richard E. Davis, Department of Biology,
            San Francisco State University, 1600 Holloway Ave, San Francisco,
            CA 94132, USA
FEATURES             Location/Qualifiers
     source          1..439
                     /organism="Fasciola hepatica"
                     /db_xref="taxon:6192"
                     /clone="Fas 3-4"
                     /clone_lib="Lambda gt11 cDNA library"
     misc_RNA        1..28
                     /note="spliced leader; 28 nucleotides of 37 nucleotide
                     spliced leader sequence"
     misc_feature    1..>439
                     /note="trans-spliced mRNA"
     polyA_site      439
                     /note="9 A nucleodites"
BASE COUNT      128 a     81 c     89 g    141 t
ORIGIN
```

```
              GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      June 13, 2001, 19:59:02 ; Search time 31.01 Seconds
                                        (without alignments)
                                        45.224 Million cell updates/sec

Title:          US-09-331-930A-2
Perfect score:  394
Sequence:       1 MIEVVCNDRLGKKVRVKCNT.........HVSLGDYEIHDGMNLELYYQ 73

Scoring table:  BLOSUM62
Gapop 10.0 , Gapext 0.5

rched:       185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters:      185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Issued_Patents_AA:*
         1:  /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
         2:  /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
         3:  /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
         4:  /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
         5:  /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
         6:  /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result NO.	Score	Query Match	Match Length	DB	ID	Description
1	60	15.2	494	3	US-08-889-841-19	Sequence 19, Appl
2	59	15.0	455	3	US-08-889-841-46	Sequence 46, Appl
3	59	15.0	850	2	US-08-448-603A-28	Sequence 28, Appl
4	59	15.0	850	3	US-09-134-075-28	Sequence 28, Appl
5	58.5	14.8	498	3	US-08-889-841-2	Sequence 2, Appli
6	58.5	14.8	498	3	US-08-889-841-5	Sequence 5, Appli
7	58	14.7	483	3	US-08-889-841-31	Sequence 31, Appl
8	58	14.7	487	3	US-08-889-841-33	Sequence 33, Appl
9	57.5	14.6	855	1	US-08-022-835-6	Sequence 6, Appli
10	57.5	14.6	855	2	US-08-388-809-6	Sequence 6, Appli
11	57.5	14.6	855	1	US-08-647-714-6	Sequence 6, Appli
12	56.5	14.3	222	1	US-08-240-496A-2	Sequence 2, Appli
13	56.5	14.3	507	2	US-08-037-816A-16	Sequence 16, Appl
14	56.5	14.3	507	2	US-08-037-816A-28	Sequence 28, Appl
15	56.5	14.3	507	2	US-08-530-146-16	Sequence 16, Appl
16	56.5	14.3	507	2	US-08-530-146-28	Sequence 28, Appl
17	56.5	14.3	606	2	US-08-392-806A-6	Sequence 6, Appli
18	56	14.2	270	4	US-08-878-474-1	Sequence 1, Appli
19	56	14.2	469	3	US-08-889-841-23	Sequence 23, Appl
20	55.5	14.1	665	3	US-08-448-603A-32	Sequence 32, Appl
21	55.5	14.1	665	3	US-09-134-075-32	Sequence 32, Appl
22	54	13.7	474	3	US-08-889-841-36	Sequence 36, Appl
23	54	13.7	474	3	US-08-889-841-39	Sequence 39, Appl
24	53.5	13.6	658	1	US-08-409-995-5	Sequence 5, Appli
25	53.5	13.6	658	3	US-08-685-467-5	Sequence 5, Appli
26	53.5	13.6	658	4	US-08-913-942-5	Sequence 5, Appli
27	53.5	13.6	666	1	US-08-441-139-16	Sequence 16, Appl
28	53.5	13.6	1098	1	US-08-409-995-2	Sequence 2, Appli
29	53.5	13.6	1098	3	US-08-685-467-2	Sequence 2, Appli
30	53.5	13.6	1098	4	US-09-377-155-32	Sequence 32, Appl
31	53.5	13.6	1098	4	US-08-913-942-2	Sequence 2, Appli
32	53.5	13.6	1724	1	US-08-325-547-3	Sequence 3, Appli
33	53	13.5	84	4	US-09-042-785A-30	Sequence 30, Appl
34	53	13.5	84	4	US-09-042-785A-31	Sequence 31, Appl
35	53	13.5	305	4	US-09-190-821-2	Sequence 2, Appli
36	53	13.5	573	4	US-09-042-785A-2	Sequence 2, Appli
37	53	13.5	605	4	US-09-042-785A-23	Sequence 23, Appl
38	53	13.5	655	3	US-08-959-382-2	Sequence 2, Appli
39	52.5	13.3	654	3	US-08-392-828C-2	Sequence 2, Appli
40	52.5	13.3	654	3	US-09-330-945-2	Sequence 2, Appli
41	51.5	13.1	496	3	US-08-889-841-12	Sequence 12, Appl
42	51.5	13.1	496	3	US-08-889-841-16	Sequence 16, Appl
43	51.5	13.1	654	1	US-08-441-139-11	Sequence 11, Appl
44	51.5	13.1	855	2	US-08-482-090-12	Sequence 12, Appl
45	51.5	13.1	856	2	US-08-481-700B-8	Sequence 8, Appli

ALIGNMENTS

```
RESULT  1
US-08-889-841-19
; Sequence 19, Application US/08889841B
; Patent No. 6090392
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
; FILE REFERENCE: 14918-703CIP
; CURRENT APPLICATION NUMBER: US/08/889,841B
; CURRENT FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: US 60/676,737
; EARLIER FILING DATE: 1996-07-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 19
;   LENGTH: 494
;   TYPE: PRT
;   ORGANISM: HIV
US-08-889-841-19

Query Match             15.2%;  Score 60;  DB 3;  Length 494;
Best Local Similarity   29.9%;  Pred. NO. 4.6;
Matches    23;  Conservative   12;  Mismatches   26;  Indels  16;  Gaps   4;

Qy     2 IEVVC------NDRLGKKV--RVKCNTDDTIGDLKKLIAAQTGTRWNKIV-----LKKW 47
         :!|:  |   !!!:|   | !!|:  :|!|::!  :!:||:|  :!:!||   ||:
Db   259 VEINCIRPNNNTRKGIHIGPGRAWYATGEIVGDIRKAYCNISRTKWNNTLIQIANKLKEK 318

Qy    48 Y--TIFKDHVSLGDYEI 62
         |  : ||!:!||! ||
Db   319 YNTTISFNRSSGGDPEI 335


RESULT  2
US-08-889-841-46
; Sequence 46, Application US/08889841B
; Patent No. 6090392
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
; FILE REFERENCE: 14918-703CIP
; CURRENT APPLICATION NUMBER: US/08/889,841B
; CURRENT FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: US 60/676,737
; EARLIER FILING DATE: 1996-07-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 46
;   LENGTH: 455
```

```
; TYPE: PRT
; ORGANISM: HIV
US-08-889-841-46

Query Match              15.0%;   Score 59;  DB 3;  Length 455;
Best Local Similarity  23.2%;  Pred. No. 5.7;
Matches  19;  Conservative  14;  Mismatches  21;  Indels  28;  Gaps  3;

Qy    9 RLGKKVRVKCN------------TDDTIGDLKKLIAAQTGTRWNKIV------ 43
          : |: : |:           : |||::  : |:||
Db  245 QLNESVEINCTRPNNTRRSIHIGPGRAFYATGEIIGDIRQAHCNLSSTKWNNTLKQIVT 304

Qy   44 -LKKWY--TIFKDHVSLGDYEI 62
         |::: || :| |||| ||
Db  305 KLREHFNKTIVFNHSSGGDPEI 326

RESULT 3
US-08-448-603A-28
; Sequence 28, Application US/08448603A
; Patent No. 5864027
; GENERAL INFORMATION:
;   APPLICANT: Berman, Phillip W.
;   APPLICANT: Nakamura, Gerald R.
;   TITLE OF INVENTION: HIV Envelope Polypeptides
;   NUMBER OF SEQUENCES: 33
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: McCutchen, Doyle, Brown & Enersen, LLP
;     STREET: 3 Embarcadero Center
;     CITY: San Francisco
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 94111
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/448,603A
;     FILING DATE: 07-JUN-1994
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/072,833
;     FILING DATE: 07-JUN-93
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Halliday, Emily
;     REGISTRATION NUMBER: 38903
;     REFERENCE/DOCKET NUMBER: 14918-704
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 415-393-2000
;     TELEFAX: 415-393-2286
;     TELEX:
; INFORMATION FOR SEQ ID NO: 28:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 850 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
US-08-448-603A-28

Query Match              15.0%;   Score 59;  DB 2;  Length 850;
Best Local Similarity  23.2%;  Pred. No. 13;
Matches  19;  Conservative  14;  Mismatches  21;  Indels  28;  Gaps  3;

Qy    9 RLGKKVRVKCN------------TDDTIGDLKKLIAAQTGTRWNKIV------ 43
          : |: : |:           : |||::  : |:||
Db  285 QLNESVEINCTRPNNTRRSIHIGPGRAFYATGEIIGDIRQAHCNLSSTKWNNTLKQIVT 344

Qy   44 -LKKWY--TIFKDHVSLGDYEI 62
         |::: || :| |||| ||
```

```
Db  345 KLREHFNKTIVFNHSSGGDPEI 366

RESULT 4
US-09-134-075-28
; Sequence 28, Application US/09134075
; Patent No. 6042836
; GENERAL INFORMATION:
;   APPLICANT: Berman, Phillip W.
;   APPLICANT: Nakamura, Gerald R.
;   TITLE OF INVENTION:  HIV Envelope Polypeptides
;   NUMBER OF SEQUENCES: 33
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: McCutchen, Doyle, Brown & Enersen, LLP
;     STREET: 3 Embarcadero Center
;     CITY: San Francisco
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 94111
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/134,075
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/448,603
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Halliday, Emily
;     REGISTRATION NUMBER: 38903
;     REFERENCE/DOCKET NUMBER: 14918-704
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 415-393-2000
;     TELEFAX: 415-393-2286
;     TELEX:
; INFORMATION FOR SEQ ID NO: 28:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 850 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
US-09-134-075-28

Query Match              15.0%;   Score 59;  DB 3;  Length 850;
Best Local Similarity  23.2%;  Pred. No. 13;
Matches  19;  Conservative  14;  Mismatches  21;  Indels  28;  Gaps  3;

Qy    9 RLGKKVRVKCN------------TDDTIGDLKKLIAAQTGTRWNKIV------ 43
          : |: : |:           : |||::  : |:||
Db  285 QLNESVEINCTRPNNTRRSIHIGPGRAFYATGEIIGDIRQAHCNLSSTKWNNTLKQIVT 344

Qy   44 -LKKWY--TIFKDHVSLGDYEI 62
         |::: || :| |||| ||
Db  345 KLREHFNKTIVFNHSSGGDPEI 366

RESULT 5
US-08-889-841-2
; Sequence 2, Application US/08889841B
; Patent No. 6090392
; GENERAL INFORMATION:
;   APPLICANT: Berman, Phillip W.
;   TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
;   FILE REFERENCE: 14918-703CIP
;   CURRENT APPLICATION NUMBER: US/08/889,841B
;   CURRENT FILING DATE: 1997-07-08
;   EARLIER APPLICATION NUMBER: US 60/676,737
;   EARLIER FILING DATE: 1996-07-08
```

; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 498
; TYPE: PRT
; ORGANISM: HIV
US-08-889-841-2

```
Query Match         14.8%;  Score 58.5; DB 3;  Length 498;
Best Local Similarity  36.5%;  Pred. No. 7.5;
Matches  19; Conservative  7; Mismatches  17; Indels  9; Gaps  2;

Qy  20 TDDTIGDLKLIAAQTGTRWN---KIVLKKW------YTIFKDHVSLGDYEI 62
        | | ||::: :| :|| | |:||    :| :|-|| || ||
Db 298 TGDIRGDIRQAHCNISGAKWNNTLKKVVKKLKEQFPNKTIVFNHSSGGDPEI 349
```

ULT 6
08-889-841-5
; Sequence 5, Application US/08889841B
; Patent No. 6090392
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
; FILE REFERENCE: 14918-703CIP
; CURRENT APPLICATION NUMBER: US/08/889,841B
; CURRENT FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: US 60/676,737
; EARLIER FILING DATE: 1996-07-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 498
; TYPE: PRT
; ORGANISM: HIV
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)..(498)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-08-889-841-5

```
Query Match         14.8%;  Score 58.5; DB 3;  Length 498;
Best Local Similarity  34.6%;  Pred. No. 7.5;
Matches  18; Conservative  8; Mismatches  17; Indels  9; Gaps  2;

Qy  20 TDDTIGDLKLIAAQTGTRWNK------IVLKKWY---TIFKDHVSLGDYEI 62
        | | ||::: :| :||      || :| :|| || ||
Db 298 TGDIRGDIRQAHCNISGAKWNNTLKKVVIKLKEQFPNKTIVFNHSSGGDPEI 349
```

RESULT 7
US-08-889-841-31
; Sequence 31, Application US/08889841B
; Patent No. 6090392
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
; FILE REFERENCE: 14918-703CIP
; CURRENT APPLICATION NUMBER: US/08/889,841B
; CURRENT FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: US 60/676,737
; EARLIER FILING DATE: 1996-07-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 483
; TYPE: PRT
; ORGANISM: HIV
US-08-889-841-31

```
Query Match         14.7%;  Score 58;  DB 3;  Length 483;
Best Local Similarity  27.5%;  Pred. No. 8.4;
Matches  22; Conservative  13; Mismatches  25; Indels  20; Gaps  4;

Qy   2 IEVVCNDRLGKKVRVKCN-------TDDTIGDLKKLIAAQTGTRWNKIV-----LKK 46
         ::: | ||| |||     | | |||||   | :| |
Db 255 VKINCT-RLGNNTRKSINIGPGRVLYATGEIIGDIRQAHCNISRAQWNKTLEKVVDKLRK 313

Qy  47 WY----TIFKDHVSLGDYEI 62
         || : | ||| ||
Db 314 QFGDNTTIAFNRSSGGDPEI 333
```

RESULT 8
US-08-889-841-33
; Sequence 33, Application US/08889841B
; Patent No. 6090392
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
; FILE REFERENCE: 14918-703CIP
; CURRENT APPLICATION NUMBER: US/08/889,841B
; CURRENT FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: US 60/676,737
; EARLIER FILING DATE: 1996-07-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 487
; TYPE: PRT
; ORGANISM: HIV
US-08-889-841-33

```
Query Match         14.7%;  Score 58;  DB 3;  Length 487;
Best Local Similarity  27.5%;  Pred. No. 8.5;
Matches  22; Conservative  13; Mismatches  25; Indels  20; Gaps  4;

Qy   2 IEVVCNDRLGKKVRVKCN-------TDDTIGDLKKLIAAQTGTRWNKIV-----LKK 46
         ::: | ||| |||     | | |||||   | :|| |
Db 255 VKINCT-RLGNNTRKSINIGPGRVLYATGEIIGDIRQAHCNISRAQWNKTLEKVVDKLRK 313

Qy  47 WY----TIFKDHVSLGDYEI 62
         || : | ||| ||
Db 314 QFGDNTTIAFNRSSGGDPEI 333
```

RESULT 9
US-08-022-835-6
; Sequence 6, Application US/08022835
; Patent No. 5420030
; GENERAL INFORMATION:
; APPLICANT: Reitz Jr., Marvin S.
; APPLICANT: Franchini, Genoveffa
; APPLICANT: Markham, Phillip D.
; APPLICANT: Gallo, Robert C.
; APPLICANT: Lori, Franco C.
; APPLICANT: Popovic, Mikulas
; APPLICANT: Garnter, Suzanne
; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1 AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: Eleventh Floor, 1615 L. Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```
   SOFTWARE:  PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
   APPLICATION NUMBER:  US/08/022,835
   FILING DATE:
   CLASSIFICATION:  435
PRIOR APPLICATION DATA:
   APPLICATION NUMBER:  US 07/599,491
   FILING DATE:  17-OCT-1990
ATTORNEY/AGENT INFORMATION:
   NAME:  Scott, Watson T.
   REGISTRATION NUMBER:  26,581
   REFERENCE/DOCKET NUMBER:  WTS/5683/78507/SRL
TELECOMMUNICATION INFORMATION:
   TELEPHONE:  (202) 861-3000
   TELEFAX:  (202) 822-0944
   TELEX:  6714627 CUSH
INFORMATION FOR SEQ ID NO: 6:
   SEQUENCE CHARACTERISTICS:
   LENGTH:  855 amino acids
   TYPE:  amino acid
   TOPOLOGY:  linear
   MOLECULE TYPE:  protein
08-022-835-6

Query Match          14.6%;  Score 57.5;  DB 1;  Length 855;
Best Local Similarity  29.5%;  Pred. No. 20;
Matches  23;  Conservative  11;  Mismatches  27;  Indels  17;  Gaps  4;

Qy    2 IEVVC---NDRLGKKV-----RVKCNTDDTIGDLKKLIAAQTGTRW----NKIVLKKW-- 47
         :|:|: |: :|:   |::||:::   : :|     |||:|
Db  296 VEINCTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHCNLSRAKWNDTLNKIVIKLREQ 355

Qy   48 ---YTIFKDHVSLGDYEI 62
          || |||| || ||
Db  356 FGNKTIVFKHSSGGDPEI 373

RESULT 10
US-08-388-809-6
; Sequence 6, Application US/08388809
; Patent No. 5576000
; GENERAL INFORMATION:
   APPLICANT:  REITZ, JR., MARVIN S.; FRANCHINI,
   APPLICANT:  GENOVEFFA; MARKHAM, PHILIP D. GALLO, ROBERT
   APPLICANT:  C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND
   APPLICANT:  GARTNER, SUZANNE
   TITLE OF INVENTION:  MOLECULAR CLONES OF HIV-1
   TITLE OF INVENTION:  AND USES THEREOF
   NUMBER OF SEQUENCES:  6
CORRESPONDENCE ADDRESS:
   ADDRESSEE:  MORGAN & FINNEGAN
   STREET:  345 PARK AVENUE
   CITY:  NEW YORK
   STATE:  NEW YORK
   COUNTRY:  USA
   ZIP:  10154
COMPUTER READABLE FORM:
   MEDIUM TYPE:  FLOPPY DISK, 3.5"
   COMPUTER:  IBM PC COMPATIBLE
   OPERATING SYSTEM:  PC-DOS/MS-DOS
   SOFTWARE:  WORDPERFECT 5.1
CURRENT APPLICATION DATA:
   APPLICATION NUMBER:  US/08/388,809
   FILING DATE:  15-FEB-1995
PRIOR APPLICATION DATA:
   APPLICATION NUMBER:  US 08/022,835
   FILING DATE:  25-FEB-1993
   APPLICATION NUMBER:  US 07/599,491
   FILING DATE:  17-OCT-1990
ATTORNEY/AGENT INFORMATION:
   NAME:  LESLIE A. SERUNIAN
   REGISTRATION NUMBER:  35,353
   REFERENCE/DOCKET NUMBER:  2026-4092US2
TELECOMMUNICATION INFORMATION:
   TELEPHONE:  (212) 758-4800
   TELEFAX:  (212) 751-6849
   TELEX:  421792
INFORMATION FOR SEQ ID NO: 6:
   SEQUENCE CHARACTERISTICS:
   LENGTH:  855 amino acids
   TYPE:  amino acid
   TOPOLOGY:  linear
   MOLECULE TYPE:  protein
US-08-388-809-6

Query Match          14.6%;  Score 57.5;  DB 1;  Length 855;
Best Local Similarity  29.5%;  Pred. No. 20;
Matches  23;  Conservative  11;  Mismatches  27;  Indels  17;  Gaps  4;

Qy    2 IEVVC---NDRLGKKV-----RVKCNTDDTIGDLKKLIAAQTGTRW----NKIVLKKW-- 47
         :|:|: |: :|:   |::||:::   : :|     |||:|
Db  296 VEINCTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHCNLSRAKWNDTLNKIVIKLREQ 355

Qy   48 ---YTIFKDHVSLGDYEI 62
          || |||| || ||
Db  356 FGNKTIVFKHSSGGDPEI 373

RESULT 11
US-08-647-714-6
; Sequence 6, Application US/08647714
; Patent No. 5869313
; GENERAL INFORMATION:
   APPLICANT:  REITZ, JR., MARVIN S.; FRANCHINI,
   APPLICANT:  GENOVEFFA; MARKHAM, PHILIP D. GALLO, ROBERT
   APPLICANT:  C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND
   APPLICANT:  GARTNER, SUZANNE
   TITLE OF INVENTION:  MOLECULAR CLONES OF HIV-1
   TITLE OF INVENTION:  AND USES THEREOF
   NUMBER OF SEQUENCES:  6
CORRESPONDENCE ADDRESS:
   ADDRESSEE:  MORGAN & FINNEGAN
   STREET:  345 PARK AVENUE
   CITY:  NEW YORK
   STATE:  NEW YORK
   COUNTRY:  USA
   ZIP:  10154
COMPUTER READABLE FORM:
   MEDIUM TYPE:  FLOPPY DISK, 3.5"
   COMPUTER:  IBM PC COMPATIBLE
   OPERATING SYSTEM:  PC-DOS/MS-DOS
   SOFTWARE:  WORDPERFECT 5.1
CURRENT APPLICATION DATA:
   APPLICATION NUMBER:  US/08/647,714
   FILING DATE:
PRIOR APPLICATION DATA:
   APPLICATION NUMBER:  US/08/388,809
   FILING DATE:  15-FEB-1995
   APPLICATION NUMBER:  US 08/022,835
   FILING DATE:  25-FEB-1993
   APPLICATION NUMBER:  US 07/599,491
   FILING DATE:  17-OCT-1990
ATTORNEY/AGENT INFORMATION:
   NAME:  LESLIE A. SERUNIAN
   REGISTRATION NUMBER:  35,353
   REFERENCE/DOCKET NUMBER:  2026-4092US2
TELECOMMUNICATION INFORMATION:
   TELEPHONE:  (212) 758-4800
   TELEFAX:  (212) 751-6849
   TELEX:  421792
INFORMATION FOR SEQ ID NO: 6:
   SEQUENCE CHARACTERISTICS:
   LENGTH:  855 amino acids
   TYPE:  amino acid
```

```
;    TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-647-714-6

Query Match          14.6%;   Score 57.5; DB 2;  Length 855;
Best Local Similarity  29.5%;  Pred. No. 20;
Matches  23;  Conservative  11;  Mismatches  27;  Indels  17;  Gaps  4;

Qy   2  IEVVC---NDRLGKKV-----RVKCNTDDTIGDLKKLIAAQTGTRW----NKIVLKKW-- 47
        :|:|   |: :|   |  :| |||:::  : :|   ||||:|
Db 296  VEINCTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHCNLSRAKWNDTLNKIVIKLREQ 355

Qy  48  --YTIFKDHVSLGDYEI 62
          || :| ||||| ||
Db 356  FGNKTIVFKHSSGGDPEI 373

RESULT 12
US-08-240-496A-2
; Sequence 2, Application US/08240496A
; Patent No. 5480805
; GENERAL INFORMATION:
;    APPLICANT: Wolf, Fred R.
;    APPLICANT: Cuellar, Richard E.
;    TITLE OF INVENTION: Composition for Modulating Sterols in
;    TITLE OF INVENTION: Yeast
;    NUMBER OF SEQUENCES: 2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Joanne M. Giesser
;      STREET: Amoco Corp., Suite 600, 55 Shuman Blvd.
;      CITY: Naperville
;      STATE: IL
;      COUNTRY: USA
;      ZIP: 60563
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/240,496A
;      FILING DATE: 10-MAY-1994
;      CLASSIFICATION: 435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 07/929,764
;      FILING DATE: 12-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Giesser, Joanne M.
;      REGISTRATION NUMBER: 32,838
;      REFERENCE/DOCKET NUMBER: 31,456
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (708) 717-2443
;      TELEFAX: (708) 717-2430
; INFORMATION FOR SEQ ID NO: 2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 222 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;    MOLECULE TYPE: protein
;    HYPOTHETICAL: NO
;    ANTI-SENSE: NO
US-08-240-496A-2

Query Match          14.3%;   Score 56.5;  DB 1;  Length 222;
Best Local Similarity  22.5%;  Pred. No. 5;
Matches  16;  Conservative  18;  Mismatches  22;  Indels  15;  Gaps  4;

Qy   5  VCNDRLGKKVRVK-CNTDDTIGDLKKLIAAQTGTRW-NKIVLKKWYTIFKD-------- 53
        :|: :| :   :|:|:: :|:: | |:  | :|:  |  |: ||| :|:|
Db  39  ICNSVISKHNAAEGLSTEDLLQDVRDALASHYGDEYINRYVKEEW--VFNNAGGAMGQMI 96


Qy  54  --HVSLGDYEI 62
          |: :| :| |
Db  97  ILHASVSEYLI 107

RESULT 13
US-08-037-816A-16
; Sequence 16, Application US/08037816A
; Patent No. 5869624
; GENERAL INFORMATION:
;    APPLICANT: Hasel, Karl W.
;    APPLICANT: Maddon, Paul J.
;    TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
;    TITLE OF INVENTION: THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES
;    TITLE OF INVENTION: THEREOF
;    NUMBER OF SEQUENCES: 29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Cooper & Dunham
;      STREET: 30 Rockefeller Plaza
;      CITY: New York
;      STATE: New York
;      COUNTRY: USA
;      ZIP: 10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/037,816A
;      FILING DATE: 26-MAR-1993
;      CLASSIFICATION: 435
;    ATTORNEY/AGENT INFORMATION:
;      NAME: White, John P.
;      REGISTRATION NUMBER: 28,678
;      REFERENCE/DOCKET NUMBER: 41190/JPW/AJM
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (212) 977-9550
;      TELEFAX: (212) 664-0525
;      TELEX: 422523 COOPUI
; INFORMATION FOR SEQ ID NO: 16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 507 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-037-816A-16

Query Match          14.3%;   Score 56.5;  DB 2;  Length 507;
Best Local Similarity  28.2%;  Pred. No. 14;
Matches  22;  Conservative  14;  Mismatches  25;  Indels  17;  Gaps  4;

Qy   2  IEVVC----NDRLGKKV-----RVKCNTDDTIGDLKKLIAAQTGTRWN---KIVLK---- 45
        :|: |:   |: :| |       ||:::  : :|| :|  :| :|:|
Db 295  VEINCTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHCNISRAKWNDTLKQIVIKLREQ 354

Qy  46  -KWYTIFKDHVSLGDYEI 62
         :  || :| |||| ||
Db 355  FENKTIVFNHSSGGDPEI 372

RESULT 14
US-08-037-816A-28
; Sequence 28, Application US/08037816A
; Patent No. 5869624
; GENERAL INFORMATION:
;    APPLICANT: Hasel, Karl W.
;    APPLICANT: Maddon, Paul J.
;    TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
;    TITLE OF INVENTION: THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES
;    TITLE OF INVENTION: THEREOF
;    NUMBER OF SEQUENCES: 29
```

```
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US/08/037,816
;       FILING DATE: 26-MAR-1993
;   ATTORNEY/AGENT INFORMATION:
;       NAME: White, John P.
;       REGISTRATION NUMBER: 28,678
;       REFERENCE/DOCKET NUMBER: 41190/JPW/AJM
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (212) 977-9550
;       TELEFAX: (212) 664-0525
;       TELEX: 422523 COOPUI
;   INFORMATION FOR SEQ ID NO: 16:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH: 507 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-530-146-16

Query Match       14.3%;  Score 56.5;  DB 2;  Length 507;
Best Local Similarity  28.2%;  Pred. No. 14;
Matches  22;  Conservative  14;  Mismatches  25;  Indels  17;  Gaps  4;

Qy    2 IEVVC---NDRLGKKV----RVKCNTDDTIGDLKKLIAAQTGTRWN---KIVLK---- 45
         :|:||:  |:   |: |    |   :|  |||:::   :||   :||:|
Db  295 VEINCTRPNNTRKSIHIGPGRAFYTTGEIIGDIRQAHCNISRAKWNDTLKQIVIKLREQ 354

Qy   46 -KWYTIFKDHVSLGDYEI 62
          :|| :||||:|| ||
Db  355 FENKTIVFNHSSGGDPEI 372

Search completed: June 13, 2001, 20:26:08
Job time: 1626 sec
```

```
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Cooper & Dunham
;       STREET: 30 Rockefeller Plaza
;       CITY: New York
;       STATE: New York
;       COUNTRY: USA
;       ZIP: 10112
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.24
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/037,816A
;       FILING DATE: 26-MAR-1993
;       CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;       NAME: White, John P.
;       REGISTRATION NUMBER: 28,678
;       REFERENCE/DOCKET NUMBER: 41190/JPW/AJM
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (212) 977-9550
;       TELEFAX: (212) 664-0525
;       TELEX: 422523 COOPUI
;   INFORMATION FOR SEQ ID NO: 28:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH: 507 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-037-816A-28

Query Match       14.3%;  Score 56.5;  DB 2;  Length 507;
Best Local Similarity  28.2%;  Pred. No. 14;
Matches  22;  Conservative  14;  Mismatches  25;  Indels  17;  Gaps  4;

Qy    2 IEVVC---NDRLGKKV----RVKCNTDDTIGDLKKLIAAQTGTRWN---KIVLK---- 45
         :|:||:  |:   |: |    |   :|  |||:::   :||   :||:|
Db  295 VEINCTRPNNTRKSIHIGPGRAFYTTGEIIGDIRQAHCNISRAKWNDTLKQIVIKLREQ 354

Qy   46 -KWYTIFKDHVSLGDYEI 62
          :|| :||||:|| ||
Db  355 FENKTIVFNHSSGGDPEI 372

RESULT 15
US-08-530-146-16
; Sequence 16, Application US/08530146
; Patent No. 5886163
; GENERAL INFORMATION:
;   APPLICANT: Hasel, Karl W.
;   APPLICANT: Maddon, Paul J.
;   TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
;   TITLE OF INVENTION: THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES
;   TITLE OF INVENTION: THEREOF
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Cooper & Dunham
;       STREET: 30 Rockefeller Plaza
;       CITY: New York
;       STATE: New York
;       COUNTRY: USA
;       ZIP: 10112
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.24
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/530,146
;       FILING DATE:
;       CLASSIFICATION:
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    June 13, 2001, 19:31:27 ; Search time 59.54 Seconds
                                      (without alignments)
                                      70.086 Million cell updates/sec

Title:             US-09-331-930A-2
Perfect score:     394
Sequence:          1 MIEVVCNDRLGKKVRVKCNT........HVSLGDYEIHDGMNLELYYQ 73

Scoring table:     BLOSUM62
                   Gapop 10.0 , Gapext 0.5

  arched:          390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters:      390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  A_Geneseq_0401:*
     1: /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
     2: /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
     3: /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
     4: /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
     5: /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
     6: /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
     7: /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
     8: /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
     9: /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
    10: /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
    11: /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
    12: /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
    13: /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
    14: /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
    15: /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
    16: /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
    17: /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
    18: /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
    19: /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
    20: /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
    21: /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*
    22: /SIDS6/gcgdata/geneseq/geneseqp/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	394	100.0	73	20	Y08413	P. obesus beacon p
2	394	100.0	73	21	B36290	Israeli sand rat b
3	313	79.4	73	21	G41925	Arabidopsis thalia
4	313	79.4	96	21	G41924	Arabidopsis thalia
5	312	79.2	73	21	G40885	Zea mays protein f
6	300	76.1	73	20	G27281	Zea mays protein f
7	163	41.4	33	20	Y08414	Human beacon prote
8	147.5	37.4	32	21	B36291	Human beacon (shor
9	80	20.3	477	21	G38562	Arabidopsis thalia
10	80	20.3	553	21	G38561	Arabidopsis thalia
11	80	20.3	631	21	G38560	Arabidopsis thalia
12	67	17.0	236	21	B41138	Human ORFX ORF902
13	60	15.2	289	21	B42562	Human ORFX ORF2326
14	60	15.2	295	21	G23780	Arabidopsis thalia
15	60	15.2	296	21	G23779	Arabidopsis thalia
16	60	15.2	323	21	G23778	Arabidopsis thalia
17	60	15.2	494	19	W37060	HIV-1 breakthrough
18	59.5	15.1	513	20	W97587	Human immunodefici
19	59	15.0	455	19	W37069	HIV-1 gp120 MN-GNE
20	59	15.0	687	22	B49465	Canine retrovirus
21	59	15.0	850	16	R67724	gp120 from the HIV
22	58.5	14.8	365	19	B18259	Plasmodium falcipa
23	58.5	14.8	498	19	W37054	HIV-1 breakthrough
24	58.5	14.8	498	19	W37055	HIV-1 breakthrough
25	58.5	14.8	855	21	Y96946	HIV synthetic Env
26	58	14.7	483	19	W37064	HIV-1 breakthrough
27	58	14.7	487	19	W37065	HIV-1 breakthrough
28	58	14.7	822	19	W69847	Amino acid sequenc
29	57.5	14.6	368	21	G44390	Arabidopsis thalia
30	57.5	14.6	465	21	G44389	Arabidopsis thalia
31	57.5	14.6	498	21	G44388	Arabidopsis thalia
32	57.5	14.6	660	21	B54249	Human pancreatic c
33	57.5	14.6	662	20	Y05377	Human HCMV inducib
34	57.5	14.6	855	18	R14905	HIV-1 BA-L clone-e
35	57.5	14.6	855	15	W11581	Human Immunodefici
36	57.5	14.6	855	20	W88113	Env protein of the
37	57	14.5	97	21	G15024	Arabidopsis thalia
38	57	14.5	139	21	G15023	Arabidopsis thalia
39	56.5	14.3	132	21	G44405	Arabidopsis thalia
40	56.5	14.3	222	17	R90670	S. cerevisiae delt
41	56.5	14.3	507	15	R60783	HIV virus-1(JR-FL)
42	56.5	14.3	507	15	R60789	HIV virus-1(JR-FL)
43	56.5	14.3	849	21	Y96945	HIV synthetic Env
44	56	14.2	88	21	G15025	Arabidopsis thalia
45	56	14.2	270	19	W41250	Xenopus cerberus p

ALIGNMENTS

```
RESULT    1
Y08413
ID   Y08413 standard; Protein; 73 AA.
XX
AC   Y08413;
XX
DT   24-JUL-1999  (first entry)
XX
DE   P. obesus beacon protein.
XX
KW   Beacon; hypothalamus; obese; lean; agonist; antagonist; treatment;
KW   obesity; anorexia; weight maintenance; energy imbalance; diabetes;
KW   metabolic syndrome; dyslipidemia; hypertension; insulin resistance;
KW   medicament; livestock; diagnosis.
XX
OS   Psammomys obesus.
XX
PN   WO9923217-A1.
XX
PD   14-MAY-1999.
XX
PF   30-OCT-1998;    98WO-AU00902.
XX
PR   11-NOV-1997;    97AU-0000323.
PR   31-OCT-1997;    97AU-0000117.
XX
PA   (ITDI-) INT DIABETES INST.
PA   (UYDE-) UNIV DEAKIN.
XX
PI   Collier G,  Zimmet Pz;
XX
DR   WPI; 1999-337484/28.
DR   N-PSDB; X57359.
XX
```

```
PT   New gene encoding a beacon protein associated with modulation of
PT   obesity, diabetes and metabolic energy levels
XX
PS   Claim 2; Page 50; 85pp; English.
XX
CC   This invention describes a novel beacon protein and its encoding nucleic
CC   acid which is expressed in larger amounts in hypothalamus tissue of obese
CC   animals compared to lean animals. Agonists and antagonists of beacon can
CC   be used to treat obesity, anorexia, weight maintenance, energy imbalance,
CC   diabetes, metabolic syndrome, dyslipidemia, hypertension and/or insulin
CC   resistance. The beacon protein, itself is used to manufacture medicaments
CC   for treatment of obesity, anorexia, energy imbalance or diabetes. The
CC   treatment is contemplated for both human and animals, such as those
CC   important to the livestock industry. The antibody and polynucleotides are
CC   useful in diagnosis of conditions as above.
XX
SQ   Sequence   73 AA;

Query Match          100.0%;  Score 394;  DB 20;  Length 73;
Best Local Similarity 100.0%;  Pred. No. 1.2e-42;
Matches  73;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDY 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 mievvcndrlgkkvrvkcntddtigdlkkliaaqtgtrwnkivlkkwytifkdhvslgdy 60

Qy    61 EIHDGMNLELYYQ 73
           |||||||||||||
Db    61 eihdgmnlelyyq 73

RESULT   2
B36290
ID   B36290 standard; Protein; 73 AA.
XX
AC   B36290;
XX
DT   23-FEB-2001 (first entry)
XX
DE   Israeli sand rat beacon ligand.
XX
KW   Israeli sand rat; beacon; obesity; NIDDM; energy balance; diabetes;
KW   ligand.
XX
OS   Psammomys obesus.
XX
PN   WO20000064931-A1.
XX
PD   02-NOV-2000.
XX
PR   19-APR-2000; 2000WO-AU00342.
XX
PR   23-APR-1999;  99AU-0009919.
PR   24-MAR-2000; 2000AU-0006454.
XX
PA   (AUTO-) AUTOGEN PTY LTD.
XX
PI   Collier G, Walder K, Zimmet P;
XX
DR   WPI; 2000-687311/67.
DR   N-PSDB; C81767.
XX
PT   Ligand of beacon protein useful for treating obesity, anorexia, energy
PT   imbalance, diabetes, metabolic syndrome, dyslipidemia, hypertension and
PT   insulin resistance -
XX
PS   Claim 3; Fig 1; 67pp; English.
XX
CC   The present invention is related to the isolation of a ligand known as
CC   beacon from the Israeli sand rat. Beacon is associated with the
CC   regulation of energy balance, and the protein, its coding sequence and
CC   analogues can be used in the treatment of diabetes, obesity, anorexia,
```

```
CC   energy imbalance, metabolic syndrome, dyslipidaemia, hypertension and
CC   insulin resistance. In addition, they can be used in agriculture to
CC   produce leaner animals.
XX
SQ   Sequence   73 AA;

Query Match          100.0%;  Score 394;  DB 21;  Length 73;
Best Local Similarity 100.0%;  Pred. No. 1.2e-42;
Matches  73;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDY 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 mievvcndrlgkkvrvkcntddtigdlkkliaaqtgtrwnkivlkkwytifkdhvslgdy 60

Qy    61 EIHDGMNLELYYQ 73
           |||||||||||||
Db    61 eihdgmnlelyyq 73

RESULT   3
G41925
ID   G41925 standard; Protein; 73 AA.
XX
AC   G41925;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 52222.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
PR   23-APR-1999;  99US-0130891.
PR   28-APR-1999;  99US-0131449.
PR   30-APR-1999;  99US-0132048.
PR   30-APR-1999;  99US-0132407.
PR   04-MAY-1999;  99US-0132484.
PR   05-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   06-MAY-1999;  99US-0132487.
PR   07-MAY-1999;  99US-0132863.
PR   11-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
PR   14-MAY-1999;  99US-0134219.
PR   14-MAY-1999;  99US-0134221.
PR   14-MAY-1999;  99US-0134370.
PR   18-MAY-1999;  99US-0134768.
PR   19-MAY-1999;  99US-0134941.
PR   20-MAY-1999;  99US-0135124.
PR   21-MAY-1999;  99US-0135353.
```

```
PR   24-MAY-1999;   99US-0135629.        PR   02-AUG-1999;   99US-0146389.
PR   25-MAY-1999;   99US-0136021.        PR   03-AUG-1999;   99US-0147038.
PR   27-MAY-1999;   99US-0136392.        PR   04-AUG-1999;   99US-0147204.
PR   28-MAY-1999;   99US-0136782.        PR   04-AUG-1999;   99US-0147302.
PR   01-JUN-1999;   99US-0137222.        PR   05-AUG-1999;   99US-0147192.
PR   03-JUN-1999;   99US-0137528.        PR   05-AUG-1999;   99US-0147260.
PR   04-JUN-1999;   99US-0137502.        PR   06-AUG-1999;   99US-0147303.
PR   07-JUN-1999;   99US-0137724.        PR   06-AUG-1999;   99US-0147416.
PR   08-JUN-1999;   99US-0138094.        PR   09-AUG-1999;   99US-0147493.
PR   10-JUN-1999;   99US-0138540.        PR   09-AUG-1999;   99US-0147935.
PR   14-JUN-1999;   99US-0138847.        PR   10-AUG-1999;   99US-0148171.
PR   16-JUN-1999;   99US-0139119.        PR   11-AUG-1999;   99US-0148319.
PR   16-JUN-1999;   99US-0139452.        PR   12-AUG-1999;   99US-0148341.
PR   16-JUN-1999;   99US-0139453.        PR   13-AUG-1999;   99US-0148565.
PR   17-JUN-1999;   99US-0139492.        PR   13-AUG-1999;   99US-0148684.
PR   18-JUN-1999;   99US-0139454.        PR   16-AUG-1999;   99US-0149368.
PR   18-JUN-1999;   99US-0139455.        PR   17-AUG-1999;   99US-0149175.
PR   18-JUN-1999;   99US-0139456.        PR   18-AUG-1999;   99US-0149426.
PR   18-JUN-1999;   99US-0139457.        PR   20-AUG-1999;   99US-0149722.
PR   18-JUN-1999;   99US-0139458.        PR   20-AUG-1999;   99US-0149723.
PR   18-JUN-1999;   99US-0139459.        PR   20-AUG-1999;   99US-0149929.
PR   18-JUN-1999;   99US-0139460.        PR   23-AUG-1999;   99US-0149902.
PR   18-JUN-1999;   99US-0139461.        PR   23-AUG-1999;   99US-0149930.
PR   18-JUN-1999;   99US-0139462.        PR   25-AUG-1999;   99US-0150566.
PR   18-JUN-1999;   99US-0139463.        PR   26-AUG-1999;   99US-0150884.
PR   18-JUN-1999;   99US-0139750.        PR   27-AUG-1999;   99US-0151065.
PR   18-JUN-1999;   99US-0139763.        PR   27-AUG-1999;   99US-0151066.
PR   21-JUN-1999;   99US-0139817.        PR   27-AUG-1999;   99US-0151080.
PR   22-JUN-1999;   99US-0139899.        PR   30-AUG-1999;   99US-0151303.
PR   23-JUN-1999;   99US-0140353.        PR   31-AUG-1999;   99US-0151438.
PR   23-JUN-1999;   99US-0140354.        PR   01-SEP-1999;   99US-0151930.
PR   24-JUN-1999;   99US-0140695.        PR   07-SEP-1999;   99US-0152363.
PR   28-JUN-1999;   99US-0140823.        PR   10-SEP-1999;   99US-0153070.
PR   29-JUN-1999;   99US-0140991.        PR   13-SEP-1999;   99US-0153758.
PR   30-JUN-1999;   99US-0141287.        PR   15-SEP-1999;   99US-0154018.
PR   01-JUL-1999;   99US-0141842.        PR   16-SEP-1999;   99US-0154039.
PR   01-JUL-1999;   99US-0142154.        PR   20-SEP-1999;   99US-0154779.
PR   02-JUL-1999;   99US-0142055.        PR   22-SEP-1999;   99US-0155139.
PR   06-JUL-1999;   99US-0142390.        PR   23-SEP-1999;   99US-0155486.
PR   08-JUL-1999;   99US-0142803.        PR   24-SEP-1999;   99US-0155659.
PR   09-JUL-1999;   99US-0142920.        PR   28-SEP-1999;   99US-0156458.
PR   12-JUL-1999;   99US-0142977.        PR   29-SEP-1999;   99US-0156596.
PR   13-JUL-1999;   99US-0143542.        PR   04-OCT-1999;   99US-0157117.
PR   14-JUL-1999;   99US-0143624.        PR   05-OCT-1999;   99US-0157753.
PR   15-JUL-1999;   99US-0144005.        PR   06-OCT-1999;   99US-0157865.
PR   16-JUL-1999;   99US-0144085.        PR   07-OCT-1999;   99US-0158029.
PR   16-JUL-1999;   99US-0144086.        PR   08-OCT-1999;   99US-0158232.
PR   19-JUL-1999;   99US-0144325.        PR   12-OCT-1999;   99US-0158369.
PR   19-JUL-1999;   99US-0144331.        PR   13-OCT-1999;   99US-0159293.
PR   19-JUL-1999;   99US-0144332.        PR   13-OCT-1999;   99US-0159294.
PR   19-JUL-1999;   99US-0144333.        PR   13-OCT-1999;   99US-0159295.
PR   19-JUL-1999;   99US-0144334.        PR   14-OCT-1999;   99US-0159329.
PR   19-JUL-1999;   99US-0144335.        PR   14-OCT-1999;   99US-0159330.
PR   20-JUL-1999;   99US-0144352.        PR   14-OCT-1999;   99US-0159331.
PR   20-JUL-1999;   99US-0144632.        PR   14-OCT-1999;   99US-0159637.
PR   21-JUL-1999;   99US-0144884.        PR   18-OCT-1999;   99US-0159638.
PR   21-JUL-1999;   99US-0144814.        PR   21-OCT-1999;   99US-0159584.
PR   21-JUL-1999;   99US-0145086.        PR   21-OCT-1999;   99US-0160741.
PR   21-JUL-1999;   99US-0145088.        PR   21-OCT-1999;   99US-0160767.
PR   22-JUL-1999;   99US-0145085.        PR   22-OCT-1999;   99US-0160768.
PR   22-JUL-1999;   99US-0145087.        PR   22-OCT-1999;   99US-0160770.
PR   22-JUL-1999;   99US-0145089.        PR   25-OCT-1999;   99US-0160814.
PR   22-JUL-1999;   99US-0145192.        PR   25-OCT-1999;   99US-0160815.
PR   23-JUL-1999;   99US-0145145.        PR   26-OCT-1999;   99US-0160980.
PR   23-JUL-1999;   99US-0145218.        PR   26-OCT-1999;   99US-0160981.
PR   23-JUL-1999;   99US-0145224.        PR   28-OCT-1999;   99US-0160989.
PR   26-JUL-1999;   99US-0145276.                          99US-0161404.
PR   27-JUL-1999;   99US-0145913.                          99US-0161405.
PR   27-JUL-1999;   99US-0145918.                          99US-0161406.
PR   27-JUL-1999;   99US-0145919.                          99US-0161359.
PR   28-JUL-1999;   99US-0145951.                          99US-0161360.
PR   02-AUG-1999;   99US-0146386.                          99US-0161361.
PR   02-AUG-1999;   99US-0146388.                          99US-0161920.
```

```
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

Query Match         79.4%;   Score 313;  DB 21;  Length 73;
Best Local Similarity   80.6%;   Pred. No. 2.1e-32;
Matches    58;  Conservative   6;  Mismatches   8;  Indels   0;  Gaps   0;

Qy    1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDY 60
        |||| ||||||||| |||||||||::|||||| || ::||| :::||::| ||
Db    1 mievvlndrlgkkvrvkcndddtigdlkklvaaqtgtraekiriqkwyniykdhitlkdy 60

Qy   61 EIHDGMNLELYY 72
        ||||| |||||
Db   61 eihdgmglelyy 72


RESULT  4
G41924
ID   G41924 standard; Protein; 96 AA.
XX
AC   G41924;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 52221.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.           PR   28-MAY-1999;   99US-0136782.
PR   05-MAR-1999;   99US-0123180.           PR   01-JUN-1999;   99US-0137222.
PR   09-MAR-1999;   99US-0123548.           PR   03-JUN-1999;   99US-0137528.
PR   23-MAR-1999;   99US-0125788.           PR   04-JUN-1999;   99US-0137502.
PR   25-MAR-1999;   99US-0126264.           PR   07-JUN-1999;   99US-0137724.
PR   29-MAR-1999;   99US-0126785.           PR   08-JUN-1999;   99US-0138094.
PR   01-APR-1999;   99US-0127462.           PR   10-JUN-1999;   99US-0138540.
PR   06-APR-1999;   99US-0128234.           PR   14-JUN-1999;   99US-0138847.
PR   08-APR-1999;   99US-0128714.           PR   16-JUN-1999;   99US-0139119.
PR   16-APR-1999;   99US-0129845.           PR   16-JUN-1999;   99US-0139452.
PR   19-APR-1999;   99US-0130077.           PR   17-JUN-1999;   99US-0139453.
PR   21-APR-1999;   99US-0130449.           PR   18-JUN-1999;   99US-0139492.
PR   23-APR-1999;   99US-0130510.           PR   18-JUN-1999;   99US-0139454.
PR   23-APR-1999;   99US-0130891.           PR   18-JUN-1999;   99US-0139455.
PR   28-APR-1999;   99US-0131449.           PR   18-JUN-1999;   99US-0139456.
PR   30-APR-1999;   99US-0132048.           PR   18-JUN-1999;   99US-0139457.
PR   04-MAY-1999;   99US-0132407.           PR   18-JUN-1999;   99US-0139458.
PR   05-MAY-1999;   99US-0132484.           PR   18-JUN-1999;   99US-0139459.
PR   06-MAY-1999;   99US-0132485.           PR   18-JUN-1999;   99US-0139460.
PR   06-MAY-1999;   99US-0132486.           PR   18-JUN-1999;   99US-0139461.
PR   07-MAY-1999;   99US-0132487.           PR   18-JUN-1999;   99US-0139462.
PR   07-MAY-1999;   99US-0132863.           PR   18-JUN-1999;   99US-0139463.
PR   11-MAY-1999;   99US-0134256.           PR   18-JUN-1999;   99US-0139750.
PR   14-MAY-1999;   99US-0134218.           PR   18-JUN-1999;   99US-0139763.
PR   14-MAY-1999;   99US-0134219.           PR   21-JUN-1999;   99US-0139817.
PR   14-MAY-1999;   99US-0134221.           PR   22-JUN-1999;   99US-0139899.
PR   18-MAY-1999;   99US-0134370.           PR   23-JUN-1999;   99US-0140353.
PR   19-MAY-1999;   99US-0134768.           PR   23-JUN-1999;   99US-0140354.
PR   20-MAY-1999;   99US-0134941.           PR   24-JUN-1999;   99US-0140695.
PR   21-MAY-1999;   99US-0135124.           PR   28-JUN-1999;   99US-0140823.
PR   24-MAY-1999;   99US-0135353.           PR   29-JUN-1999;   99US-0140991.
PR   25-MAY-1999;   99US-0135629.           PR   30-JUN-1999;   99US-0141287.
PR   27-MAY-1999;   99US-0136021.           PR   01-JUL-1999;   99US-0141842.
                    99US-0136392.           PR   01-JUL-1999;   99US-0142154.
                                            PR   02-JUL-1999;   99US-0142055.
                                            PR   06-JUL-1999;   99US-0142390.
                                            PR   08-JUL-1999;   99US-0142803.
                                            PR   09-JUL-1999;   99US-0142920.
                                            PR   12-JUL-1999;   99US-0142977.
                                            PR   13-JUL-1999;   99US-0143542.
                                            PR   14-JUL-1999;   99US-0143624.
                                            PR   15-JUL-1999;   99US-0144005.
                                            PR   16-JUL-1999;   99US-0144085.
                                            PR   16-JUL-1999;   99US-0144086.
                                            PR   19-JUL-1999;   99US-0144325.
                                            PR   19-JUL-1999;   99US-0144331.
                                            PR   19-JUL-1999;   99US-0144332.
                                            PR   19-JUL-1999;   99US-0144333.
                                            PR   19-JUL-1999;   99US-0144334.
                                            PR   19-JUL-1999;   99US-0144335.
                                            PR   20-JUL-1999;   99US-0144352.
                                            PR   20-JUL-1999;   99US-0144632.
                                            PR   21-JUL-1999;   99US-0144884.
                                            PR   21-JUL-1999;   99US-0144814.
                                            PR   21-JUL-1999;   99US-0145086.
                                            PR   22-JUL-1999;   99US-0145088.
                                            PR   22-JUL-1999;   99US-0145085.
                                            PR   22-JUL-1999;   99US-0145087.
                                            PR   22-JUL-1999;   99US-0145089.
                                            PR   23-JUL-1999;   99US-0145192.
                                            PR   23-JUL-1999;   99US-0145145.
                                            PR   23-JUL-1999;   99US-0145218.
                                            PR   23-JUL-1999;   99US-0145224.
                                            PR   26-JUL-1999;   99US-0145276.
                                            PR   27-JUL-1999;   99US-0145913.
                                            PR   27-JUL-1999;   99US-0145918.
                                            PR   27-JUL-1999;   99US-0145919.
                                            PR   28-JUL-1999;   99US-0145951.
                                            PR   02-AUG-1999;   99US-0146386.
                                            PR   02-AUG-1999;   99US-0146388.
                                            PR   02-AUG-1999;   99US-0146389.
                                            PR   03-AUG-1999;   99US-0147038.
                                            PR   04-AUG-1999;   99US-0147204.
```

```
Query Match                    79.4%;  Score 313; DB 21; Length 96;
Best Local Similarity  80.6%;          Pred. No. 3e-32;
Matches  58;  Conservative   6;  Mismatches   8;  Indels   0;  Gaps   0;

Qy    1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDY 60
        ||||| |||||||||| |||||||| ||||||||||||||| ||  |  |   |  ||||::||
Db   24 mievvlndrlgkkvrvkcnddtigdlkklvaaqtgtraekiriqkwyniykdhitlkdy 83

Qy   61 EIHDGMNLELY 72
        |||||| |||||
Db   84 eihdgmqlelyy 95


RESULT  5
G40885
ID   G40885 standard; Protein; 73 AA.
XX
AC   G40885;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Zea mays protein fragment SEQ ID NO: 50790.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence; corn.
XX
OS   Zea mays subsp. mays.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135355.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
```

```
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   18-OCT-1999;   99US-0159638.
PR   21-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   22-OCT-1999;   99US-0160814.
PR   22-OCT-1999;   99US-0160815.
PR   25-OCT-1999;   99US-0160980.
PR   25-OCT-1999;   99US-0160981.
PR   26-OCT-1999;   99US-0160989.
PR   26-OCT-1999;   99US-0161404.
PR   26-OCT-1999;   99US-0161405.
PR   28-OCT-1999;   99US-0161406.
PR   28-OCT-1999;   99US-0161359.
PR   28-OCT-1999;   99US-0161360.
PR   29-OCT-1999;   99US-0161361.
                    99US-0161920.
                    99US-0161992.
                    99US-0161993.
                    99US-0162142.
```

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 03-AUG-1999; 99US-0146388.
PR 04-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 11-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0155596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 18-OCT-1999; 99US-0159637.
PR 21-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 22-OCT-1999; 99US-0160770.
PR 22-OCT-1999; 99US-0160814.
PR 25-OCT-1999; 99US-0160815.
PR 25-OCT-1999; 99US-0160980.
PR 25-OCT-1999; 99US-0160981.
PR 26-OCT-1999; 99US-0160989.
PR 26-OCT-1999; 99US-0161404.
PR 28-OCT-1999; 99US-0161405.
PR 28-OCT-1999; 99US-0161406.
PR 29-OCT-1999; 99US-0161359.
PR ; 99US-0161360.
PR ; 99US-0161361.
PR ; 99US-0161920.
PR ; 99US-0161992.
PR ; 99US-0161993.
PR ; 99US-0162142.

Query Match   79.2%;   Score 312; DB 21; Length 73;
Best Local Similarity 80.6%;   Pred. No. 2.8e-32;

```
Matches   58;  Conservative   6;  Mismatches   8;  Indels   0;  Gaps   0;

Qy      1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDY 60
          |||| |||||||||| |||||||||| ||||||| || ::||| |:||::| ||
Db      1 mievvlndrlgkkvrvkcneddtigdlkklvaaqtgtrpekiriqkwyniykdhitlkdy 60

Qy     61 EIHDGMNLELYY 72
          | ||||| |||||
Db     61 eihdgmglelyy 72


RESULT   6
G27281
ID   G27281 standard; Protein; 73 AA.
XX
AC   G27281;
XX
DT   17-OCT-2000 (first entry)
DE   Zea mays protein fragment SEQ ID NO: 32055.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence; corn.
XX
OS   Zea mays subsp. mays.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.        PR   10-JUN-1999; 99US-0138540.
PR   05-MAR-1999; 99US-0123180.        PR   10-JUN-1999; 99US-0138847.
PR   09-MAR-1999; 99US-0123548.        PR   14-JUN-1999; 99US-0139119.
PR   23-MAR-1999; 99US-0125788.        PR   16-JUN-1999; 99US-0139452.
PR   25-MAR-1999; 99US-0126264.        PR   16-JUN-1999; 99US-0139453.
PR   29-MAR-1999; 99US-0126785.        PR   17-JUN-1999; 99US-0139492.
PR   01-APR-1999; 99US-0127462.        PR   18-JUN-1999; 99US-0139454.
PR   06-APR-1999; 99US-0128234.        PR   18-JUN-1999; 99US-0139455.
PR   08-APR-1999; 99US-0128714.        PR   18-JUN-1999; 99US-0139456.
PR   16-APR-1999; 99US-0129845.        PR   18-JUN-1999; 99US-0139457.
PR   19-APR-1999; 99US-0130077.        PR   18-JUN-1999; 99US-0139458.
PR   21-APR-1999; 99US-0130449.        PR   18-JUN-1999; 99US-0139459.
PR   23-APR-1999; 99US-0130510.        PR   18-JUN-1999; 99US-0139460.
PR   23-APR-1999; 99US-0130891.        PR   18-JUN-1999; 99US-0139461.
PR   28-APR-1999; 99US-0131449.        PR   18-JUN-1999; 99US-0139462.
PR   30-APR-1999; 99US-0132048.        PR   18-JUN-1999; 99US-0139463.
PR   30-APR-1999; 99US-0132407.        PR   18-JUN-1999; 99US-0139750.
PR   04-MAY-1999; 99US-0132484.        PR   21-JUN-1999; 99US-0139763.
PR   05-MAY-1999; 99US-0132485.        PR   22-JUN-1999; 99US-0139817.
PR   06-MAY-1999; 99US-0132486.        PR   22-JUN-1999; 99US-0139899.
PR   07-MAY-1999; 99US-0132487.        PR   23-JUN-1999; 99US-0140353.
PR   11-MAY-1999; 99US-0132863.        PR   23-JUN-1999; 99US-0140354.
PR   14-MAY-1999; 99US-0134256.        PR   24-JUN-1999; 99US-0140695.
PR   14-MAY-1999; 99US-0134218.        PR   28-JUN-1999; 99US-0140823.
PR   14-MAY-1999; 99US-0134219.        PR   29-JUN-1999; 99US-0140991.
PR   14-MAY-1999; 99US-0134221.        PR   30-JUN-1999; 99US-0141287.
PR   14-MAY-1999; 99US-0134370.        PR   01-JUL-1999; 99US-0141842.
PR   18-MAY-1999; 99US-0134768.        PR   01-JUL-1999; 99US-0142154.
PR   19-MAY-1999; 99US-0134941.        PR   02-JUL-1999; 99US-0142055.
PR   20-MAY-1999; 99US-0135124.        PR   06-JUL-1999; 99US-0142390.
PR   21-MAY-1999; 99US-0135353.        PR   08-JUL-1999; 99US-0142803.
PR   24-MAY-1999; 99US-0135629.        PR   09-JUL-1999; 99US-0142920.
PR   25-MAY-1999; 99US-0136021.        PR   12-JUL-1999; 99US-0142977.
PR   27-MAY-1999; 99US-0136392.        PR   13-JUL-1999; 99US-0143542.
PR   28-MAY-1999; 99US-0136782.        PR   14-JUL-1999; 99US-0143624.
PR   01-JUN-1999; 99US-0137222.        PR   15-JUL-1999; 99US-0144005.
PR   03-JUN-1999; 99US-0137528.        PR   16-JUL-1999; 99US-0144085.
PR   04-JUN-1999; 99US-0137502.        PR   16-JUL-1999; 99US-0144086.
PR   07-JUN-1999; 99US-0137724.        PR   19-JUL-1999; 99US-0144325.
PR   08-JUN-1999; 99US-0138094.        PR   19-JUL-1999; 99US-0144331.
                                       PR   19-JUL-1999; 99US-0144332.
                                       PR   19-JUL-1999; 99US-0144333.
                                       PR   19-JUL-1999; 99US-0144334.
                                       PR   20-JUL-1999; 99US-0144335.
                                       PR   20-JUL-1999; 99US-0144352.
                                       PR   20-JUL-1999; 99US-0144632.
                                       PR   21-JUL-1999; 99US-0144884.
                                       PR   21-JUL-1999; 99US-0144814.
                                       PR   21-JUL-1999; 99US-0145086.
                                       PR   22-JUL-1999; 99US-0145088.
                                       PR   22-JUL-1999; 99US-0145085.
                                       PR   22-JUL-1999; 99US-0145087.
                                       PR   23-JUL-1999; 99US-0145089.
                                       PR   23-JUL-1999; 99US-0145192.
                                       PR   23-JUL-1999; 99US-0145145.
                                       PR   26-JUL-1999; 99US-0145218.
                                       PR   27-JUL-1999; 99US-0145224.
                                       PR   27-JUL-1999; 99US-0145276.
                                       PR   28-JUL-1999; 99US-0145913.
                                       PR   02-AUG-1999; 99US-0145918.
                                       PR   02-AUG-1999; 99US-0145919.
                                       PR   03-AUG-1999; 99US-0145951.
                                       PR   04-AUG-1999; 99US-0146386.
                                       PR   05-AUG-1999; 99US-0146388.
                                       PR   05-AUG-1999; 99US-0146389.
                                       PR   06-AUG-1999; 99US-0147038.
                                       PR   06-AUG-1999; 99US-0147204.
                                       PR   09-AUG-1999; 99US-0147302.
                                       PR               ; 99US-0147192.
                                       PR               ; 99US-0147260.
                                       PR               ; 99US-0147303.
                                       PR               ; 99US-0147416.
                                       PR               ; 99US-0147493.
```

```
        |||||  ||||||||||| |||||| ||||||||||| ||:||||||  ||  :::|||  |:|:||| ||
Db     1 mievvlndrlgkkvrvkcneddtigdlnklvaaqtgtrpekirigkwyniykdyitlkdy 60

Qy    61 EIHDGMNLELYY 72
           |:|||| |||||
Db    61 evhdgmglelyy 72


RESULT    7
Y08414
ID   Y08414 standard; Protein; 33 AA.
XX
AC   Y08414;
XX
DT   24-JUL-1999  (first entry)
XX
DE   Human beacon protein.
XX
KW   Beacon; hypothalamus; obese; lean; agonist; antagonist; treatment;
KW   obesity; anorexia; weight maintenance; energy imbalance; diabetes;
KW   metabolic syndrome; dyslipidemia; hypertension; insulin resistance;
KW   medicament; livestock; diagnosis; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 15
FT                   /label= unknown
FT                   /note= "encoded by CNC"
XX
PN   WO9923217-A1.
XX
PD   14-MAY-1999.
XX
PF   30-OCT-1998;      98WO-AU00902.
XX
PR   11-NOV-1997;      97AU-0000323.
PR   31-OCT-1997;      97AU-0000117.
XX
PA   (ITDI-) INT DIABETES INST.
PA   (UVDE-) UNIV DEAKIN.
XX
PI   Collier G,  Zimmet PZ;
XX
DR   WPI; 1999-337484/28.
DR   N-PSDB; X57370.
XX
PT   New gene encoding a beacon protein associated with modulation of
PT   obesity, diabetes and metabolic energy levels
XX
PS   Claim 2; Page 54; 85pp; English.
XX
CC   This invention describes a novel beacon protein and its encoding nucleic
CC   acid which is expressed in larger amounts in hypothalamus tissue of obese
CC   animals compared to lean animals. Agonists and antagonists of beacon can
CC   be used to treat obesity, anorexia, weight maintenance, energy imbalance,
CC   diabetes, metabolic syndrome, dyslipidemia, hypertension and/or insulin
CC   resistance. The beacon protein, itself is used to manufacture medicaments
CC   for treatment of obesity, anorexia, energy imbalance or diabetes. The
CC   treatment is contemplated for both human and animals, such as those
CC   important to the livestock industry. The antibody and polynucleotides are
CC   useful in diagnosis of conditions as above.
XX
SQ   Sequence   33 AA;

Query Match             41.4%;  Score 163;  DB 20;  Length 33;
Best Local Similarity   97.0%;           Pred. No. 7.7e-14;
Matches   32;  Conservative    0;  Mismatches    1;   Indels    0;  Gaps    0;

Qy     1 MIEVVCNDRLGKKVRVKCNTDDFIGDLKKLIAA 33
           ||||||||||||| |||||||||||||||||||
Db     1 mievvcndrlgkkvxvkcntddtigdlkkliaa 33
```

```
PR   09-AUG-1999;  99US-0147935.
PR   10-AUG-1999;  99US-0148171.
PR   11-AUG-1999;  99US-0148319.
PR   12-AUG-1999;  99US-0148341.
PR   13-AUG-1999;  99US-0148565.
PR   13-AUG-1999;  99US-0148684.
PR   16-AUG-1999;  99US-0149368.
PR   17-AUG-1999;  99US-0149175.
PR   18-AUG-1999;  99US-0149426.
PR   18-AUG-1999;  99US-0149722.
PR   20-AUG-1999;  99US-0149723.
PR   20-AUG-1999;  99US-0149929.
PR   23-AUG-1999;  99US-0149902.
PR   23-AUG-1999;  99US-0149930.
PR   25-AUG-1999;  99US-0150566.
PR   26-AUG-1999;  99US-0150884.
PR   27-AUG-1999;  99US-0151065.
PR   27-AUG-1999;  99US-0151066.
PR   27-AUG-1999;  99US-0151080.
PR   30-AUG-1999;  99US-0151303.
PR   31-AUG-1999;  99US-0151438.
PR   01-SEP-1999;  99US-0151930.
PR   07-SEP-1999;  99US-0152363.
PR   10-SEP-1999;  99US-0153070.
PR   13-SEP-1999;  99US-0153758.
PR   15-SEP-1999;  99US-0154018.
PR   16-SEP-1999;  99US-0154039.
PR   20-SEP-1999;  99US-0154779.
PR   22-SEP-1999;  99US-0155138.
PR   23-SEP-1999;  99US-0155486.
PR   24-SEP-1999;  99US-0155659.
PR   28-SEP-1999;  99US-0156458.
PR   29-SEP-1999;  99US-0156596.
PR   04-OCT-1999;  99US-0157117.
PR   05-OCT-1999;  99US-0157753.
PR   06-OCT-1999;  99US-0157865.
PR   07-OCT-1999;  99US-0158029.
PR   08-OCT-1999;  99US-0158232.
PR   12-OCT-1999;  99US-0158369.
PR   13-OCT-1999;  99US-0159293.
PR   13-OCT-1999;  99US-0159294.
PR   14-OCT-1999;  99US-0159295.
PR   14-OCT-1999;  99US-0159329.
PR   14-OCT-1999;  99US-0159330.
PR   14-OCT-1999;  99US-0159331.
PR   14-OCT-1999;  99US-0159637.
PR   14-OCT-1999;  99US-0159638.
PR   18-OCT-1999;  99US-0159584.
PR   21-OCT-1999;  99US-0160741.
PR   21-OCT-1999;  99US-0160767.
PR   21-OCT-1999;  99US-0160768.
PR   21-OCT-1999;  99US-0160770.
PR   21-OCT-1999;  99US-0160814.
PR   21-OCT-1999;  99US-0160815.
PR   22-OCT-1999;  99US-0160980.
PR   22-OCT-1999;  99US-0160981.
PR   22-OCT-1999;  99US-0160989.
PR   25-OCT-1999;  99US-0161404.
PR   25-OCT-1999;  99US-0161405.
PR   25-OCT-1999;  99US-0161406.
PR   26-OCT-1999;  99US-0161359.
PR   26-OCT-1999;  99US-0161360.
PR   26-OCT-1999;  99US-0161361.
PR   28-OCT-1999;  99US-0161920.
PR   28-OCT-1999;  99US-0161992.
PR   28-OCT-1999;  99US-0161993.
PR   29-OCT-1999;  99US-0162142.


Query Match             76.1%;  Score 300;  DB 21;  Length 73;
Best Local Similarity   76.4%;           Pred. No. 9.4e-31;
Matches   55;  Conservative    8;  Mismatches    9;   Indels    0;  Gaps    0;

Qy     1 MIEVVCNDRLGKKVRVKCNTDDFIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDY 60
```

```
RESULT   8
B36291
ID   B36291 standard; Protein; 32 AA.
XX
AC   B36291;
XX
DT   23-FEB-2001  (first entry)
XX
DE   Human beacon (short form) protein sequence.
XX
KW   Israeli sand rat; beacon; obesity; NIDDM; energy balance; diabetes;
KW   ligand; human.
XX
OS   Homo sapiens.
XX
XX
   WO20064931-A1.
XX
   02-NOV-2000.
XX
PF   19-APR-2000; 2000WO-AU00342.
XX
PR   23-APR-1999;   99AU-0009919.
PR   24-MAR-2000; 2000AU-0006454.
XX
PA   (AUTO-) AUTOGEN PTY LTD.
XX
PI   Collier G,  Walder K,  Zimmet P;
XX
DR   WPI: 2000-687311/67.
DR   N-PSDB; C81770.
XX
PT   Ligand of beacon protein useful for treating obesity, anorexia, energy
PT   imbalance, diabetes, metabolic syndrome, dyslipidemia, hypertension and
PT   insulin resistance -
XX
PS   Disclosure; Fig 1B; 67pp; English.
XX
CC   The present invention is related to the isolation of a ligand known as
CC   beacon from the Israeli sand rat. Beacon is associated with the
CC   regulation of energy balance, and the protein, its coding sequence and
CC   analogues can be used in the treatment of diabetes, obesity, anorexia,
CC   energy imbalance, metabolic syndrome, dyslipidaemia, hypertension and
CC   insulin resistance. In addition, they can be used in agriculture to
CC   produce leaner animals.
XX
SQ   Sequence    32 AA;

Query Match             37.4%;    Score 147.5; DB 21;  Length 32;
Best Local Similarity   93.9%;    Pred. No. 6.8e-12;
Matches   31;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy    1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAA 33
        ||||||||||||| |||||||||||| |||||
Db    1 mievvcndrlgkkvxvkcntddtigdl-kliaa 32


RESULT   9
G38562
ID   G38562 standard; Protein; 477 AA.
XX
AC   G38562;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 47590.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
```

```
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;  2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
PR   23-APR-1999;  99US-0130891.
PR   28-APR-1999;  99US-0131449.
PR   30-APR-1999;  99US-0132048.
PR   30-APR-1999;  99US-0132407.
PR   04-MAY-1999;  99US-0132484.
PR   05-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   06-MAY-1999;  99US-0132487.
PR   07-MAY-1999;  99US-0132863.
PR   11-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
PR   14-MAY-1999;  99US-0134219.
PR   14-MAY-1999;  99US-0134221.
PR   14-MAY-1999;  99US-0134370.
PR   18-MAY-1999;  99US-0134768.
PR   19-MAY-1999;  99US-0134941.
PR   20-MAY-1999;  99US-0135124.
PR   21-MAY-1999;  99US-0135353.
PR   24-MAY-1999;  99US-0135629.
PR   25-MAY-1999;  99US-0136021.
PR   27-MAY-1999;  99US-0136392.
PR   28-MAY-1999;  99US-0136782.
PR   01-JUN-1999;  99US-0137222.
PR   03-JUN-1999;  99US-0137528.
PR   04-JUN-1999;  99US-0137502.
PR   07-JUN-1999;  99US-0137724.
PR   08-JUN-1999;  99US-0138094.
PR   10-JUN-1999;  99US-0138540.
PR   10-JUN-1999;  99US-0138847.
PR   14-JUN-1999;  99US-0139119.
PR   16-JUN-1999;  99US-0139452.
PR   16-JUN-1999;  99US-0139453.
PR   17-JUN-1999;  99US-0139492.
PR   18-JUN-1999;  99US-0139454.
PR   18-JUN-1999;  99US-0139455.
PR   18-JUN-1999;  99US-0139456.
PR   18-JUN-1999;  99US-0139457.
PR   18-JUN-1999;  99US-0139458.
PR   18-JUN-1999;  99US-0139459.
PR   18-JUN-1999;  99US-0139460.
PR   18-JUN-1999;  99US-0139461.
PR   18-JUN-1999;  99US-0139462.
PR   18-JUN-1999;  99US-0139463.
PR   18-JUN-1999;  99US-0139750.
PR   18-JUN-1999;  99US-0139763.
PR   21-JUN-1999;  99US-0139817.
PR   22-JUN-1999;  99US-0139899.
PR   23-JUN-1999;  99US-0140353.
PR   23-JUN-1999;  99US-0140354.
PR   24-JUN-1999;  99US-0140695.
PR   28-JUN-1999;  99US-0140823.
PR   29-JUN-1999;  99US-0140991.
```

```
PR   30-JUN-1999;   99US-0141287.        PR   15-SEP-1999;   99US-0154018.
PR   01-JUL-1999;   99US-0141842.        PR   16-SEP-1999;   99US-0154039.
PR   01-JUL-1999;   99US-0142154.        PR   20-SEP-1999;   99US-0154779.
PR   02-JUL-1999;   99US-0142055.        PR   22-SEP-1999;   99US-0155139.
PR   06-JUL-1999;   99US-0142390.        PR   23-SEP-1999;   99US-0155486.
PR   08-JUL-1999;   99US-0142803.        PR   24-SEP-1999;   99US-0155659.
PR   09-JUL-1999;   99US-0142920.        PR   28-SEP-1999;   99US-0155458.
PR   12-JUL-1999;   99US-0142977.        PR   29-SEP-1999;   99US-0156596.
PR   13-JUL-1999;   99US-0143542.        PR   04-OCT-1999;   99US-0157117.
PR   14-JUL-1999;   99US-0143624.        PR   05-OCT-1999;   99US-0157753.
PR   15-JUL-1999;   99US-0144005.        PR   06-OCT-1999;   99US-0157865.
PR   16-JUL-1999;   99US-0144085.        PR   07-OCT-1999;   99US-0158029.
PR   16-JUL-1999;   99US-0144086.        PR   08-OCT-1999;   99US-0158232.
PR   19-JUL-1999;   99US-0144325.        PR   12-OCT-1999;   99US-0158369.
PR   19-JUL-1999;   99US-0144331.        PR   13-OCT-1999;   99US-0159293.
PR   19-JUL-1999;   99US-0144332.        PR   13-OCT-1999;   99US-0159294.
PR   19-JUL-1999;   99US-0144333.        PR   13-OCT-1999;   99US-0159295.
PR   19-JUL-1999;   99US-0144334.        PR   14-OCT-1999;   99US-0159329.
PR   20-JUL-1999;   99US-0144335.        PR   14-OCT-1999;   99US-0159330.
PR   20-JUL-1999;   99US-0144352.        PR   14-OCT-1999;   99US-0159331.
PR   20-JUL-1999;   99US-0144632.        PR   14-OCT-1999;   99US-0159637.
PR   21-JUL-1999;   99US-0144884.        PR   14-OCT-1999;   99US-0159638.
PR   21-JUL-1999;   99US-0144814.        PR   18-OCT-1999;   99US-0159584.
PR   21-JUL-1999;   99US-0145086.        PR   21-OCT-1999;   99US-0160741.
PR   21-JUL-1999;   99US-0145088.        PR   21-OCT-1999;   99US-0160767.
PR   22-JUL-1999;   99US-0145085.        PR   21-OCT-1999;   99US-0160768.
PR   22-JUL-1999;   99US-0145087.        PR   21-OCT-1999;   99US-0160770.
PR   22-JUL-1999;   99US-0145089.        PR   21-OCT-1999;   99US-0160814.
PR   22-JUL-1999;   99US-0145192.        PR   21-OCT-1999;   99US-0160815.
PR   23-JUL-1999;   99US-0145145.        PR   22-OCT-1999;   99US-0160980.
PR   23-JUL-1999;   99US-0145218.        PR   22-OCT-1999;   99US-0160981.
PR   23-JUL-1999;   99US-0145224.        PR   22-OCT-1999;   99US-0160989.
PR   26-JUL-1999;   99US-0145276.        PR   25-OCT-1999;   99US-0161404.
PR   27-JUL-1999;   99US-0145913.        PR   25-OCT-1999;   99US-0161405.
PR   27-JUL-1999;   99US-0145918.        PR   25-OCT-1999;   99US-0161406.
PR   27-JUL-1999;   99US-0145919.        PR   26-OCT-1999;   99US-0161359.
PR   28-JUL-1999;   99US-0145951.        PR   26-OCT-1999;   99US-0161360.
PR   02-AUG-1999;   99US-0146386.        PR   26-OCT-1999;   99US-0161361.
PR   02-AUG-1999;   99US-0146388.        PR   28-OCT-1999;   99US-0161920.
PR   02-AUG-1999;   99US-0146389.        PR   28-OCT-1999;   99US-0161992.
PR   03-AUG-1999;   99US-0147038.        PR   28-OCT-1999;   99US-0161993.
PR   04-AUG-1999;   99US-0147204.        PR   29-OCT-1999;   99US-0162142.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
```

```
Query Match          20.3%;    Score 80; DB 21; Length 477;
Best Local Similarity 31.7%;   Pred. No. 0.073;
Matches  19; Conservative  16; Mismatches  25; Indels  0; Gaps  0;

Qy   11 GKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDYEIHDGMNLEL 70
        ||  .:|:  .:||| .:|| .:|  .::| ::    .:|.:|||:|||  |
Db  407 gktilevessdtianvkekiqvkegikpdqqmliffgqqledgvtlgdydihkkstlyl 466


RESULT 10
G38561
ID   G38561 standard; Protein; 553 AA.
XX
AC   G38561;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 47589.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP103405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
```

us-09-331-930a-2.rag

```
XX  25-FEB-1999;  99US-0121825.        PR  12-JUL-1999;  99US-0142977.
PR  05-MAR-1999;  99US-0123180.        PR  13-JUL-1999;  99US-0143542.
PR  09-MAR-1999;  99US-0123548.        PR  14-JUL-1999;  99US-0143624.
PR  23-MAR-1999;  99US-0125788.        PR  15-JUL-1999;  99US-0144005.
PR  25-MAR-1999;  99US-0126264.        PR  16-JUL-1999;  99US-0144085.
PR  29-MAR-1999;  99US-0126785.        PR  16-JUL-1999;  99US-0144086.
PR  01-APR-1999;  99US-0127462.        PR  19-JUL-1999;  99US-0144325.
PR  06-APR-1999;  99US-0128234.        PR  19-JUL-1999;  99US-0144331.
PR  08-APR-1999;  99US-0128714.        PR  19-JUL-1999;  99US-0144332.
PR  16-APR-1999;  99US-0129845.        PR  19-JUL-1999;  99US-0144333.
PR  19-APR-1999;  99US-0130077.        PR  19-JUL-1999;  99US-0144334.
PR  21-APR-1999;  99US-0130449.        PR  19-JUL-1999;  99US-0144335.
PR  23-APR-1999;  99US-0130510.        PR  20-JUL-1999;  99US-0144352.
PR  28-APR-1999;  99US-0130891.        PR  20-JUL-1999;  99US-0144632.
PR  30-APR-1999;  99US-0131449.        PR  21-JUL-1999;  99US-0144884.
PR  30-APR-1999;  99US-0132048.        PR  21-JUL-1999;  99US-0144814.
PR  04-MAY-1999;  99US-0132407.        PR  21-JUL-1999;  99US-0145086.
PR  05-MAY-1999;  99US-0132484.        PR  22-JUL-1999;  99US-0145008.
PR  06-MAY-1999;  99US-0132485.        PR  22-JUL-1999;  99US-0145085.
PR  06-MAY-1999;  99US-0132486.        PR  22-JUL-1999;  99US-0145087.
PR  07-MAY-1999;  99US-0132487.        PR  23-JUL-1999;  99US-0145089.
PR  11-MAY-1999;  99US-0132663.        PR  23-JUL-1999;  99US-0145192.
PR  14-MAY-1999;  99US-0134256.        PR  23-JUL-1999;  99US-0145145.
PR  14-MAY-1999;  99US-0134218.        PR  26-JUL-1999;  99US-0145218.
PR  14-MAY-1999;  99US-0134219.        PR  27-JUL-1999;  99US-0145224.
PR  14-MAY-1999;  99US-0134221.        PR  27-JUL-1999;  99US-0145276.
PR  18-MAY-1999;  99US-0134370.        PR  28-JUL-1999;  99US-0145913.
PR  19-MAY-1999;  99US-0134768.        PR  02-AUG-1999;  99US-0145918.
PR  20-MAY-1999;  99US-0134941.        PR  02-AUG-1999;  99US-0145919.
PR  21-MAY-1999;  99US-0135124.        PR  03-AUG-1999;  99US-0145951.
PR  24-MAY-1999;  99US-0135353.        PR  04-AUG-1999;  99US-0146386.
PR  25-MAY-1999;  99US-0135629.        PR  04-AUG-1999;  99US-0146388.
PR  27-MAY-1999;  99US-0136021.        PR  05-AUG-1999;  99US-0146389.
PR  28-MAY-1999;  99US-0136392.        PR  05-AUG-1999;  99US-0147038.
PR  01-JUN-1999;  99US-0137222.        PR  06-AUG-1999;  99US-0147204.
PR  03-JUN-1999;  99US-0137528.        PR  09-AUG-1999;  99US-0147302.
PR  04-JUN-1999;  99US-0137502.        PR  09-AUG-1999;  99US-0147192.
PR  07-JUN-1999;  99US-0137724.        PR  10-AUG-1999;  99US-0147260.
PR  08-JUN-1999;  99US-0138094.        PR  11-AUG-1999;  99US-0147303.
PR  10-JUN-1999;  99US-0138540.        PR  12-AUG-1999;  99US-0147416.
PR  10-JUN-1999;  99US-0138847.        PR  13-AUG-1999;  99US-0147493.
PR  14-JUN-1999;  99US-0139119.        PR  13-AUG-1999;  99US-0147935.
PR  16-JUN-1999;  99US-0139452.        PR  16-AUG-1999;  99US-0148171.
PR  16-JUN-1999;  99US-0139453.        PR  17-AUG-1999;  99US-0148319.
PR  17-JUN-1999;  99US-0139492.        PR  18-AUG-1999;  99US-0148341.
PR  18-JUN-1999;  99US-0139454.        PR  20-AUG-1999;  99US-0148565.
PR  18-JUN-1999;  99US-0139455.        PR  20-AUG-1999;  99US-0148684.
PR  18-JUN-1999;  99US-0139456.        PR  23-AUG-1999;  99US-0149368.
PR  18-JUN-1999;  99US-0139457.        PR  23-AUG-1999;  99US-0149175.
PR  18-JUN-1999;  99US-0139458.        PR  25-AUG-1999;  99US-0149426.
PR  18-JUN-1999;  99US-0139459.        PR  26-AUG-1999;  99US-0149722.
PR  18-JUN-1999;  99US-0139460.        PR  27-AUG-1999;  99US-0149723.
PR  18-JUN-1999;  99US-0139461.        PR  27-AUG-1999;  99US-0149929.
PR  18-JUN-1999;  99US-0139462.        PR  30-AUG-1999;  99US-0149902.
PR  18-JUN-1999;  99US-0139463.        PR  31-AUG-1999;  99US-0149930.
PR  18-JUN-1999;  99US-0139750.        PR  01-SEP-1999;  99US-0150566.
PR  18-JUN-1999;  99US-0139763.        PR  07-SEP-1999;  99US-0150884.
PR  21-JUN-1999;  99US-0139817.        PR  10-SEP-1999;  99US-0151065.
PR  22-JUN-1999;  99US-0139899.        PR  13-SEP-1999;  99US-0151066.
PR  23-JUN-1999;  99US-0140353.        PR  15-SEP-1999;  99US-0151080.
PR  23-JUN-1999;  99US-0140354.        PR  16-SEP-1999;  99US-0151303.
PR  24-JUN-1999;  99US-0140695.        PR  20-SEP-1999;  99US-0151438.
PR  28-JUN-1999;  99US-0140823.        PR  22-SEP-1999;  99US-0151930.
PR  29-JUN-1999;  99US-0140991.        PR  23-SEP-1999;  99US-0152363.
PR  30-JUN-1999;  99US-0141287.        PR  24-SEP-1999;  99US-0153070.
PR  01-JUL-1999;  99US-0141842.        PR  28-SEP-1999;  99US-0153758.
PR  02-JUL-1999;  99US-0142154.                          99US-0154018.
PR  06-JUL-1999;  99US-0142055.                          99US-0154039.
PR  08-JUL-1999;  99US-0142390.                          99US-0154779.
PR  08-JUL-1999;  99US-0142803.                          99US-0155139.
PR  09-JUL-1999;  99US-0142920.                          99US-0155486.
                                                         99US-0155659.
                                                         99US-0156458.
```

PR	29-SEP-1999;	99US-0155596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	28-APR-1999;	99US-0130891.
PR	30-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	04-MAY-1999;	99US-0132407.
PR	05-MAY-1999;	99US-0132484.
PR	06-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	18-MAY-1999;	99US-0134370.
PR	19-MAY-1999;	99US-0134768.
PR	20-MAY-1999;	99US-0134941.
PR	21-MAY-1999;	99US-0135124.
PR	24-MAY-1999;	99US-0135353.
PR	25-MAY-1999;	99US-0135629.
PR	27-MAY-1999;	99US-0136021.
PR	28-MAY-1999;	99US-0136392.
PR	01-JUN-1999;	99US-0136782.
PR	03-JUN-1999;	99US-0137222.
PR	04-JUN-1999;	99US-0137528.
PR	07-JUN-1999;	99US-0137502.
PR	08-JUN-1999;	99US-0137724.
PR	10-JUN-1999;	99US-0138094.
PR	14-JUN-1999;	99US-0138540.
PR	16-JUN-1999;	99US-0138847.
PR	16-JUN-1999;	99US-0139119.
PR	17-JUN-1999;	99US-0139452.
PR	18-JUN-1999;	99US-0139453.
PR	18-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	21-JUN-1999;	99US-0139763.
PR	22-JUN-1999;	99US-0139817.
PR	23-JUN-1999;	99US-0139899.
PR	24-JUN-1999;	99US-0140353.
PR	28-JUN-1999;	99US-0140354.
PR	30-JUN-1999;	99US-0140695.
PR	01-JUL-1999;	99US-0140823.
PR	01-JUL-1999;	99US-0140991.
PR	02-JUL-1999;	99US-0141287.
PR	06-JUL-1999;	99US-0141842.
PR	08-JUL-1999;	99US-0142154.
PR	09-JUL-1999;	99US-0142055.
PR	12-JUL-1999;	99US-0142390.
PR	13-JUL-1999;	99US-0142803.
PR	14-JUL-1999;	99US-0142920.
PR	15-JUL-1999;	99US-0142977.
PR	16-JUL-1999;	99US-0143542.
PR	16-JUL-1999;	99US-0143624.
PR	19-JUL-1999;	99US-0144005.
		99US-0144085.
		99US-0144086.
		99US-0144325.

Query Match   20.3%;   Score 80;   DB 21;   Length 553;
Best Local Similarity   31.7%;   Pred. No. 0.088;
Matches   19;   Conservative   16;   Mismatches   25;   Indels   0;   Gaps   0;

```
Qy     11 GKKVRVKCNTDDFTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDYEIHDGMNLEL 70
          ||  :: : ||| ::|: |  : : :|  :||:|||:|||||||||
Db    483 gktiilevessdtianvkekiqvkegikpdqqmliffgqqledgvtlgdydihkkstlyl 542
```

RESULT 11
G38560
ID   G38560 standard; Protein; 631 AA.
XX
AC   G38560;
XX
DT   18-OCT-2000   (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 47588.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.

```
PR   19-JUL-1999;  99US-0144331.
PR   19-JUL-1999;  99US-0144332.
PR   19-JUL-1999;  99US-0144333.
PR   19-JUL-1999;  99US-0144334.
PR   19-JUL-1999;  99US-0144335.
PR   20-JUL-1999;  99US-0144352.
PR   20-JUL-1999;  99US-0144632.
PR   20-JUL-1999;  99US-0144884.
PR   21-JUL-1999;  99US-0144814.
PR   21-JUL-1999;  99US-0145086.
PR   21-JUL-1999;  99US-0145088.
PR   22-JUL-1999;  99US-0145085.
PR   22-JUL-1999;  99US-0145087.
PR   22-JUL-1999;  99US-0145089.
PR   22-JUL-1999;  99US-0145192.
PR   23-JUL-1999;  99US-0145145.
PR   23-JUL-1999;  99US-0145218.
PR   26-JUL-1999;  99US-0145224.
PR   26-JUL-1999;  99US-0145276.
PR   27-JUL-1999;  99US-0145913.
PR   27-JUL-1999;  99US-0145918.
PR   27-JUL-1999;  99US-0145919.
PR   28-JUL-1999;  99US-0145951.
PR   02-AUG-1999;  99US-0146386.
PR   02-AUG-1999;  99US-0146388.
PR   02-AUG-1999;  99US-0146389.
PR   03-AUG-1999;  99US-0147038.
PR   04-AUG-1999;  99US-0147204.
PR   04-AUG-1999;  99US-0147302.
PR   05-AUG-1999;  99US-0147260.
PR   05-AUG-1999;  99US-0147303.
PR   06-AUG-1999;  99US-0147416.
PR   09-AUG-1999;  99US-0147493.
PR   09-AUG-1999;  99US-0147935.
PR   10-AUG-1999;  99US-0148171.
PR   11-AUG-1999;  99US-0148319.
PR   12-AUG-1999;  99US-0148341.
PR   13-AUG-1999;  99US-0148565.
PR   13-AUG-1999;  99US-0148684.
PR   16-AUG-1999;  99US-0149368.
PR   17-AUG-1999;  99US-0149175.
PR   18-AUG-1999;  99US-0149426.
PR   20-AUG-1999;  99US-0149722.
PR   20-AUG-1999;  99US-0149723.
PR   20-AUG-1999;  99US-0149929.
PR   23-AUG-1999;  99US-0149902.
PR   23-AUG-1999;  99US-0149930.
PR   25-AUG-1999;  99US-0150566.
PR   26-AUG-1999;  99US-0150884.
PR   27-AUG-1999;  99US-0151065.
PR   27-AUG-1999;  99US-0151066.
PR   27-AUG-1999;  99US-0151080.
PR   30-AUG-1999;  99US-0151303.
PR   31-AUG-1999;  99US-0151438.
PR   01-SEP-1999;  99US-0151930.
PR   07-SEP-1999;  99US-0152363.
PR   10-SEP-1999;  99US-0153070.
PR   13-SEP-1999;  99US-0153758.
PR   15-SEP-1999;  99US-0154018.
PR   16-SEP-1999;  99US-0154039.
PR   20-SEP-1999;  99US-0154779.
PR   22-SEP-1999;  99US-0155139.
PR   23-SEP-1999;  99US-0155486.
PR   24-SEP-1999;  99US-0155659.
PR   28-SEP-1999;  99US-0156458.
PR   29-SEP-1999;  99US-0156596.
PR   04-OCT-1999;  99US-0157117.
PR   05-OCT-1999;  99US-0157753.
PR   06-OCT-1999;  99US-0157865.
PR   07-OCT-1999;  99US-0158029.
PR   08-OCT-1999;  99US-0158232.
PR   12-OCT-1999;  99US-0158369.

PR   13-OCT-1999;  99US-0159293.
PR   13-OCT-1999;  99US-0159294.
PR   13-OCT-1999;  99US-0159295.
PR   14-OCT-1999;  99US-0159329.
PR   14-OCT-1999;  99US-0159330.
PR   14-OCT-1999;  99US-0159331.
PR   14-OCT-1999;  99US-0159637.
PR   14-OCT-1999;  99US-0159638.
PR   18-OCT-1999;  99US-0159584.
PR   21-OCT-1999;  99US-0160741.
PR   21-OCT-1999;  99US-0160767.
PR   21-OCT-1999;  99US-0160768.
PR   21-OCT-1999;  99US-0160770.
PR   21-OCT-1999;  99US-0160814.
PR   21-OCT-1999;  99US-0160815.
PR   22-OCT-1999;  99US-0160980.
PR   22-OCT-1999;  99US-0160981.
PR   22-OCT-1999;  99US-0160989.
PR   25-OCT-1999;  99US-0161404.
PR   25-OCT-1999;  99US-0161405.
PR   25-OCT-1999;  99US-0161406.
PR   26-OCT-1999;  99US-0161359.
PR   26-OCT-1999;  99US-0161360.
PR   26-OCT-1999;  99US-0161361.
PR   28-OCT-1999;  99US-0161920.
PR   28-OCT-1999;  99US-0161992.
PR   28-OCT-1999;  99US-0161993.
PR   29-OCT-1999;  99US-0162142.
```

```
Query Match             20.3%;  Score 80;  DB 21;  Length 631;
Best Local Similarity   31.7%;  Pred. No. 0.1;
Matches   19;  Conservative   16;  Mismatches   25;  Indels   0;  Gaps   0;

Qy    11 GKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDYEIHDGMNLEL 70
            || :: || ::|| ::| :  |: :  :| :    :|:||||:|| |
Db   561 gktiilevessdtianvkekiqvkegikpdqqmliffgqqledgvtlgdydihkkstlyl 620
```

```
RESULT 12
B41138
ID   B41138 standard; Protein; 236 AA.
XX
AC   B41138;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF902 polypeptide sequence SEQ ID NO:1804.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;  99US-0127607.
PR   02-APR-1999;  99US-0127636.
```

PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000;   2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shinkets RA, Leach M;
XX
DR   WPI; 2000-602362/57.
DR   N-PSDB; C75347.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 11; Page 1405-1406; 5507pp; English.
XX
CC   C74446 to C77606 encode the proteins given in B40237 to B43397, which
CC   represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC   anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC   cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC   dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC   antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC   sequences can be used for determining the presence of or predisposition
CC   to, or preventing or treating pathological conditions associated with an
CC   ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC   proteins in gene therapy vectors. The proteins and nucleic acids may be
CC   used to treat cancers, proliferative disorders, neurodegenerative
CC   disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC   diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC   storage, systemic lupus erythematosus, severe combined immunodeficiency
CC   (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC   disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC   cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC   enhance coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence   236 AA;

Query Match   17.0%;   Score 67; DB 21; Length 236;
Best Local Similarity   32.8%;   Pred. No. 1.3;
Matches   22;   Conservative   10;   Mismatches   31;   Indels   4;   Gaps   3;

Qy   5 VCNDRL--GKKVRVKCNTDFTGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDYEI 62
        :|| | |::| : : :|:| :|:| ::|:| |:: || : ||| : || : ||
Db   57 icnmsiewgarvgmvasdettftylkdrphaprgaqwdkav-aywrtlrtdddatfdael 115

Qy   63 H-DGMNL 68
        | || ||
Db 116 hvdasnl 122

RESULT 13
B42562
ID   B42562 standard; Protein; 289 AA.
XX
AC   B42562;
XX
DT   08-FEB-2001 (first entry)
XX
DE   Human ORFX ORF2326 polypeptide sequence SEQ ID NO:4652.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;

KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000;   2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000;   2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shinkets RA, Leach M;
XX
DR   WPI; 2000-602362/57.
DR   N-PSDB; C76771.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 11; Page 3851-3852; 5507pp; English.
XX
CC   C74446 to C77606 encode the proteins given in B40237 to B43397, which
CC   represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC   anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC   cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC   dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC   antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC   sequences can be used for determining the presence of or predisposition
CC   to, or preventing or treating pathological conditions associated with an
CC   ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC   proteins in gene therapy vectors. The proteins and nucleic acids may be
CC   used to treat cancers, proliferative disorders, neurodegenerative
CC   disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC   diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC   storage, systemic lupus erythematosus, severe combined immunodeficiency
CC   (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC   disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC   cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC   enhance coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence   289 AA;

Query Match   15.2%;   Score 60; DB 21; Length 289;
Best Local Similarity   21.4%;   Pred. No. 13;
Matches   12;   Conservative   15;   Mismatches   23;   Indels   6;   Gaps   2;

Qy   2 IEVVCNDRLGKKVRVKCNTDD---TIGDLKKLIAAQTGTRWNKIVLKKWYTIFKD 53
        :||| :|: :| : || :|| ::|| :||| | : | || ||| :|| ||
Db 213 ldiicetamgkniqagsnddseyvravyrmsemifprikmpw--lwldlwylmfke 266

RESULT 14
G23780
ID   G23780 standard; Protein; 295 AA.
XX
AC   G23780;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 27213.

XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.   PR   23-JUN-1999;   99US-0140353.
PR   05-MAR-1999;   99US-0123180.   PR   23-JUN-1999;   99US-0140354.
PR   09-MAR-1999;   99US-0123548.   PR   24-JUN-1999;   99US-0140695.
PR   23-MAR-1999;   99US-0125788.   PR   28-JUN-1999;   99US-0140823.
PR   25-MAR-1999;   99US-0126264.   PR   29-JUN-1999;   99US-0140991.
PR   29-MAR-1999;   99US-0126785.   PR   30-JUN-1999;   99US-0141287.
PR   01-APR-1999;   99US-0127462.   PR   01-JUL-1999;   99US-0141842.
PR   06-APR-1999;   99US-0128234.   PR   02-JUL-1999;   99US-0142154.
PR   08-APR-1999;   99US-0128714.   PR   02-JUL-1999;   99US-0142055.
PR   16-APR-1999;   99US-0129845.   PR   06-JUL-1999;   99US-0142390.
PR   19-APR-1999;   99US-0130077.   PR   08-JUL-1999;   99US-0142803.
PR   21-APR-1999;   99US-0130449.   PR   09-JUL-1999;   99US-0142920.
PR   23-APR-1999;   99US-0130510.   PR   12-JUL-1999;   99US-0142977.
PR   23-APR-1999;   99US-0130891.   PR   13-JUL-1999;   99US-0143542.
PR   28-APR-1999;   99US-0131449.   PR   14-JUL-1999;   99US-0143624.
PR   30-APR-1999;   99US-0132048.   PR   15-JUL-1999;   99US-0144005.
PR   30-APR-1999;   99US-0132407.   PR   16-JUL-1999;   99US-0144085.
PR   04-MAY-1999;   99US-0132484.   PR   16-JUL-1999;   99US-0144086.
PR   05-MAY-1999;   99US-0132485.   PR   19-JUL-1999;   99US-0144325.
PR   06-MAY-1999;   99US-0132486.   PR   19-JUL-1999;   99US-0144331.
PR   06-MAY-1999;   99US-0132487.   PR   19-JUL-1999;   99US-0144332.
PR   07-MAY-1999;   99US-0132863.   PR   19-JUL-1999;   99US-0144333.
PR   11-MAY-1999;   99US-0134256.   PR   19-JUL-1999;   99US-0144334.
PR   14-MAY-1999;   99US-0134218.   PR   19-JUL-1999;   99US-0144335.
PR   14-MAY-1999;   99US-0134219.   PR   20-JUL-1999;   99US-0144352.
PR   14-MAY-1999;   99US-0134221.   PR   20-JUL-1999;   99US-0144632.
PR   14-MAY-1999;   99US-0134370.   PR   20-JUL-1999;   99US-0144884.
PR   18-MAY-1999;   99US-0134768.   PR   21-JUL-1999;   99US-0144814.
PR   19-MAY-1999;   99US-0134941.   PR   21-JUL-1999;   99US-0145086.
PR   20-MAY-1999;   99US-0135124.   PR   21-JUL-1999;   99US-0145088.
PR   21-MAY-1999;   99US-0135353.   PR   22-JUL-1999;   99US-0145085.
PR   24-MAY-1999;   99US-0135629.   PR   22-JUL-1999;   99US-0145087.
PR   25-MAY-1999;   99US-0136021.   PR   22-JUL-1999;   99US-0145089.
PR   27-MAY-1999;   99US-0136392.   PR   23-JUL-1999;   99US-0145192.
PR   28-MAY-1999;   99US-0136782.   PR   23-JUL-1999;   99US-0145145.
PR   01-JUN-1999;   99US-0137222.   PR   23-JUL-1999;   99US-0145218.
PR   03-JUN-1999;   99US-0137528.   PR   23-JUL-1999;   99US-0145224.
PR   04-JUN-1999;   99US-0137502.   PR   26-JUL-1999;   99US-0145276.
PR   07-JUN-1999;   99US-0137724.   PR   27-JUL-1999;   99US-0145913.
PR   08-JUN-1999;   99US-0138094.   PR   27-JUL-1999;   99US-0145918.
PR   10-JUN-1999;   99US-0138540.   PR   27-JUL-1999;   99US-0145919.
PR   10-JUN-1999;   99US-0138847.   PR   28-JUL-1999;   99US-0145951.
PR   14-JUN-1999;   99US-0139119.   PR   02-AUG-1999;   99US-0146386.
PR   16-JUN-1999;   99US-0139452.   PR   02-AUG-1999;   99US-0146388.
PR   16-JUN-1999;   99US-0139457.   PR   02-AUG-1999;   99US-0146389.
PR   18-JUN-1999;   99US-0139453.   PR   03-AUG-1999;   99US-0147038.
PR   18-JUN-1999;   99US-0139458.   PR   04-AUG-1999;   99US-0147204.
PR   18-JUN-1999;   99US-0139492.   PR   04-AUG-1999;   99US-0147302.
PR   18-JUN-1999;   99US-0139459.   PR   05-AUG-1999;   99US-0147192.
PR   18-JUN-1999;   99US-0139460.   PR   05-AUG-1999;   99US-0147260.
PR   18-JUN-1999;   99US-0139461.   PR   06-AUG-1999;   99US-0147303.
PR   18-JUN-1999;   99US-0139462.   PR   06-AUG-1999;   99US-0147416.
PR   18-JUN-1999;   99US-0139463.   PR   09-AUG-1999;   99US-0147493.
PR   18-JUN-1999;   99US-0139750.   PR   09-AUG-1999;   99US-0147935.
PR   18-JUN-1999;   99US-0139763.   PR   10-AUG-1999;   99US-0149368.
PR   21-JUN-1999;   99US-0139817.   PR   11-AUG-1999;   99US-0148171.
PR   22-JUN-1999;   99US-0139899.   PR   12-AUG-1999;   99US-0148319.
PR   PR   13-AUG-1999;   99US-0148341.
PR   PR   13-AUG-1999;   99US-0148565.
PR   PR   16-AUG-1999;   99US-0148684.
PR   PR   17-AUG-1999;   99US-0149175.
PR   PR   18-AUG-1999;   99US-0149426.
PR   PR   20-AUG-1999;   99US-0149722.
PR   PR   20-AUG-1999;   99US-0149723.
PR   PR   23-AUG-1999;   99US-0149929.
PR   PR   23-AUG-1999;   99US-0149902.
PR   PR   25-AUG-1999;   99US-0149930.
PR   PR   26-AUG-1999;   99US-0150566.
PR   PR   27-AUG-1999;   99US-0150884.
PR   PR   27-AUG-1999;   99US-0151065.
PR   PR   27-AUG-1999;   99US-0151066.
PR   PR   30-AUG-1999;   99US-0151080.
PR   PR   99US-0151303.

```
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;   2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   18-JUN-1999;   99US-0139453.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139750.
PR   21-JUN-1999;   99US-0139763.
PR   22-JUN-1999;   99US-0139817.
PR   23-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   24-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.

PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

Query Match             15.2%;   Score 60;   DB 21;   Length 295;
Best Local Similarity   30.8%;   Pred. No. 14;
Matches  16;   Conservative  16;   Mismatches  9;   Indels  13;   Gaps  2;

Qy         11 GKKVRVK--------CNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKD 53
                 | :||||  |       | | |:|||:|| ||  : |  |:|| |:||
Db         37 girvrvkygsvhheisinsqstfgelkklisgatgvhhqdmqi-----iykd 83

RESULT 15
G23779
ID   G23779 standard; Protein; 296 AA.
XX
AC   G23779;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 27212.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
```

```
PR 01-JUL-1999; 99US-0142154.          PR 20-SEP-1999; 99US-0154779.
PR 02-JUL-1999; 99US-0142055.          PR 22-SEP-1999; 99US-0155139.
PR 06-JUL-1999; 99US-0142390.          PR 23-SEP-1999; 99US-0155486.
PR 08-JUL-1999; 99US-0142803.          PR 24-SEP-1999; 99US-0155659.
PR 09-JUL-1999; 99US-0142920.          PR 28-SEP-1999; 99US-0156458.
PR 12-JUL-1999; 99US-0142977.          PR 29-SEP-1999; 99US-0156596.
PR 13-JUL-1999; 99US-0143542.          PR 04-OCT-1999; 99US-0157117.
PR 14-JUL-1999; 99US-0143624.          PR 05-OCT-1999; 99US-0157753.
PR 15-JUL-1999; 99US-0144005.          PR 06-OCT-1999; 99US-0157865.
PR 16-JUL-1999; 99US-0144085.          PR 07-OCT-1999; 99US-0158029.
PR 16-JUL-1999; 99US-0144086.          PR 08-OCT-1999; 99US-0158232.
PR 19-JUL-1999; 99US-0144325.          PR 12-OCT-1999; 99US-0158369.
PR 19-JUL-1999; 99US-0144331.          PR 13-OCT-1999; 99US-0159293.
PR 19-JUL-1999; 99US-0144332.          PR 13-OCT-1999; 99US-0159294.
PR 19-JUL-1999; 99US-0144333.          PR 13-OCT-1999; 99US-0159295.
PR 19-JUL-1999; 99US-0144334.          PR 14-OCT-1999; 99US-0159329.
PR 19-JUL-1999; 99US-0144335.          PR 14-OCT-1999; 99US-0159330.
PR 20-JUL-1999; 99US-0144352.          PR 14-OCT-1999; 99US-0159331.
PR 20-JUL-1999; 99US-0144632.          PR 14-OCT-1999; 99US-0159637.
PR 20-JUL-1999; 99US-0144884.          PR 14-OCT-1999; 99US-0159638.
PR 21-JUL-1999; 99US-0144814.          PR 18-OCT-1999; 99US-0159584.
PR 21-JUL-1999; 99US-0145086.          PR 21-OCT-1999; 99US-0160741.
PR 21-JUL-1999; 99US-0145088.          PR 21-OCT-1999; 99US-0160767.
PR 22-JUL-1999; 99US-0145085.          PR 21-OCT-1999; 99US-0160768.
PR 22-JUL-1999; 99US-0145087.          PR 21-OCT-1999; 99US-0160770.
PR 22-JUL-1999; 99US-0145089.          PR 21-OCT-1999; 99US-0160814.
PR 22-JUL-1999; 99US-0145192.          PR 21-OCT-1999; 99US-0160815.
PR 23-JUL-1999; 99US-0145145.          PR 22-OCT-1999; 99US-0160980.
PR 23-JUL-1999; 99US-0145218.          PR 22-OCT-1999; 99US-0160981.
PR 23-JUL-1999; 99US-0145224.          PR 22-OCT-1999; 99US-0160989.
PR 26-JUL-1999; 99US-0145276.          PR 25-OCT-1999; 99US-0161404.
PR 27-JUL-1999; 99US-0145913.          PR 25-OCT-1999; 99US-0161405.
PR 27-JUL-1999; 99US-0145918.          PR 25-OCT-1999; 99US-0161406.
PR 27-JUL-1999; 99US-0145919.          PR 26-OCT-1999; 99US-0161359.
PR 28-JUL-1999; 99US-0145951.          PR 26-OCT-1999; 99US-0161360.
PR 02-AUG-1999; 99US-0146386.          PR 26-OCT-1999; 99US-0161361.
PR 02-AUG-1999; 99US-0146388.          PR 28-OCT-1999; 99US-0161920.
PR 03-AUG-1999; 99US-0146389.          PR 28-OCT-1999; 99US-0161992.
PR 03-AUG-1999; 99US-0147038.          PR 28-OCT-1999; 99US-0161993.
PR 04-AUG-1999; 99US-0147204.          PR 29-OCT-1999; 99US-0162142.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 05-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
```

```
Query Match            15.2%;    Score 60; DB 21; Length 296;
Best Local Similarity  30.8%;              Pred. No. 14;
Matches  16; Conservative   9; Mismatches 13; Indels  14; Gaps  2;

Qy    11 GKKVRVK-------CNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKD 53
         | :||||     |: |:||:::: ||  : :  :|||  |:||
Db    38 girvrvkygsvhheisinsqstfgelkkilsgatgvhhqdmqi----iykd 84


Search completed: June 13, 2001, 20:25:25
Job time: 3238 sec
```

```
              GenCore version 4.5
       Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    June 13, 2001, 20:25:33 ; Search time 60.15 seconds
                              (without alignments)
                        142.247 Million cell updates/sec

Title:          US-09-331-930A-2
Perfect score:  394
Sequence:       1 MIEVVCNDRLGKKVRVKCNT.........HVSLGDYEIHDGMNLELYYQ 73

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters:   374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      SPTREMBL_15:*
            1:  sp_archea:*
            2:  sp_bacteria:*
            3:  sp_fungi:*
            4:  sp_human:*
            5:  sp_invertebrate:*
            6:  sp_mammal:*
            7:  sp_mhc:*
            8:  sp_organelle:*
            9:  sp_phage:*
           10:  sp_plant:*
           11:  sp_rodent:*
           12:  sp_unclassified:*
           13:  sp_vertebrate:*
           14:  sp_virus:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	339	86.0	73	5	Q9V998	Q9v998 drosophila
2	331	84.0	73	5	P91302	P91302 caenorhabdi
3	306	77.7	73	10	Q9M1U1	Q9mlul arabidopsis
4	301	76.4	73	10	O94650	O94650 schizosacch
5	89	22.6	115	14	O90461	O90461 human immun
6	88.5	22.5	114	14	Q9WNQ6	Q9wnq6 human immun
7	88	22.3	203	14	Q9JBC9	Q9jbc9 human immun
8	87	22.1	115	14	Q71412	Q71412 human immun
9	87	22.1	216	14	O89553	O89553 human immun
10	86	21.8	115	14	Q9WNN3	Q9wnn3 human immun
11	86	21.8	164	14	O41876	O41876 human immun
12	85.5	21.7	114	14	Q71495	Q71495 human immun
13	85.5	21.7	114	14	O90467	O90467 human immun
14	85.5	21.7	176	14	Q9QCN8	Q9qcn8 human immun
15	85.5	21.7	209	14	Q75118	Q75118 human immun
16	85.5	21.7	417	14	Q9QCN7	Q9qcn7 human immun
17	85.5	21.7	857	14	Q9QML6	Q9qml6 human immun
18	85	21.6	115	14	Q70167	Q70167 human immun
19	85	21.6	115	14	Q70175	Q70175 human immun
20	85	21.6	115	14	O74431	human immun
21	85	21.6	115	14	O90411	human immun
22	85	21.6	115	14	Q9TW7	human immun
23	85	21.6	119	14	O77156	human immun
24	85	21.6	202	14	Q9JBC1	human immun
25	85	21.6	203	14	Q9JBB7	human immun
26	84.5	21.4	114	14	Q71437	human immun
27	84.5	21.4	114	14	P88770	human immun
28	84.5	21.4	159	14	Q9WBN6	human immun
29	84	21.3	115	14	Q9WNP7	human immun
30	84	21.3	212	14	Q9IVU1	human immun
31	83.5	21.2	101	14	Q9QSW3	human immun
32	83.5	21.2	163	14	O41868	human immun
33	83.5	21.2	163	14	O41877	human immun
34	83.5	21.2	164	14	O56899	human immun
35	83.5	21.2	455	14	Q9IZE4	human immun
36	83.5	21.2	543	14	Q9YUY7	human immun
37	83	21.1	175	14	P88226	human immun
38	83	21.1	200	14	Q9JBB2	human immun
39	83	21.1	200	14	Q9JBA7	human immun
40	83	21.1	201	14	Q9JBC2	human immun
41	83	21.1	201	14	Q9JBB4	human immun
42	83	21.1	201	14	Q9JBB1	human immun
43	83	21.1	201	14	Q9JBA9	human immun
44	83	21.1	201	14	Q9JBA6	human immun
45	83	21.1	201	14	Q9JBA5	human immun

ALIGNMENTS

```
RESULT 1
Q9V998
ID   Q9V998         PRELIMINARY;       PRT;    73 AA.
AC   Q9V998;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG3450 PROTEIN.
GN   CG3450.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
```

RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003790; AAF57398.1; -.
DR   FLYBASE; FBgn0033105; CG3450.
DR   INTERPRO; IPR000626; -.
DR   PROSITE; PS50053; UBIQUITIN_2; 1.
SQ   SEQUENCE   73 AA; 8570 MW; 0870111AC5586A70 CRC64;

Query Match   86.0%; Score 339; DB 5; Length 73;
Best Local Similarity   86.3%;
Matches   63; Conservative   3; Mismatches   7; Indels   0; Gaps   0;

Qy   1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDY 60
              ||| :|||||||||||||| |||||||||||||||||: |||||||||||||||
Db   1 MIEITCNDRLGKKVRVKCNPDDTIGDLKKLIAAQTGTKHEKIVLKKWYTIFKDPIRLSDY 60

Qy   61 EIHDGMNLELYYQ 73
              |||||||||||||
Db   61 EIHDGMNLELYYQ 73

RESULT 2
ID   P91302   PRELIMINARY;   PRT;   73 AA.
AC   P91302;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   SIMILARITY TO ARABIDOPSIS THALIANA UBIQUITIN-LIKE PROTEIN 8.
GN   F46F11.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Pauley A., Gattung S.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;

RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U88173; AAB42266.1; -.
DR   INTERPRO; IPR000626; -.
DR   PROSITE; PS50053; UBIQUITIN_2; 1.
SQ   SEQUENCE   73 AA; 8738 MW; 61CA839BBA4006A4 CRC64;

Query Match   84.0%; Score 331; DB 5; Length 73;
Best Local Similarity   80.8%; Pred. No. 7.2e-32;
Matches   59; Conservative   6; Mismatches   8; Indels   0; Gaps   0;

Qy   1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDY 60
              ||: |||||||||||:|||| |||||||||||||||| :|||||||||||::::1 ||
Db   1 MIEITVNDRLGKKVRIKCNPSDTIGDLKKLIAAQTGTRWEKIVLKKWYTIYKDHITLMDY 60

Qy   61 EIHDGMNLELYYQ 73
              |||:| |||||
Db   61 EIHEGFNFELYYQ 73

RESULT 3
Q9M1U1
ID   Q9M1U1   PRELIMINARY;   PRT;   73 AA.
AC   Q9M1U1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYPOTHETICAL 8.6 KDA PROTEIN.
GN   T14D3.120.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Jordan N., Bangert S., Wiedelmann R., Voss H., Unseld M., Mewes H.W.,
RA   Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RL   EMBL; AL138649; CAB72156.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   73 AA; 8621 MW; 5F26B037B7062E41 CRC64;

Query Match   77.8%; Score 306; DB 10; Length 73;
Best Local Similarity   79.2%; Pred. No. 6.3e-29;
Matches   57; Conservative   6; Mismatches   9; Indels   0; Gaps   0;

Qy   1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHVSLGDY 60
              |||| |||||||||||| |||||||||||:||||||: | :|:||| : :|| ||
Db   1 MIEVLNDRLGKKVRVKCNEEDTIGDLKKLVAAQTGTRPEKIRIQKWYNIYKDHIPLKDY 60

Qy   61 EIHDGMNLELYY 72
              |||||| ||||
Db   61 EIHDGMGLELYY 72

RESULT 4
O94650
ID   O94650   PRELIMINARY;   PRT;   73 AA.
AC   O94650;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   UBIQUITIN-LIKE PROTEIN C31E1.03.
GN   SPBC31E1.03.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetaceae; Schizosaccharomyces.

OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Oliver K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS A UBIQUITIN-LIKE DOMAIN.
DR   EMBL; AL049190; CAB39137.1; -.
DR   INTERPRO; IPR000626; -.
DR   PROSITE; PS50053; UBIQUITIN_2; 1.
KW   Hypothetical protein; Nuclear protein; Multigene family.
FT   DOMAIN   2   73   UBIQUITIN-LIKE.
SQ   SEQUENCE   73 AA;   8435 MW;   894BED56B01FE365 CRC64;

Query Match   76.4%; Score 301; DB 3; Length 73;
Best Local Similarity   73.6%; Pred. No. 2.5e-28;
Matches   53; Conservative 12; Mismatches   7; Indels   0; Gaps   0;

Qy   1 MIEVVCNDRLGKKKVRVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKKWTIFKDHVSLGDY 60
         ||:||||||||||||||||| |||:|| ||| |||| :|||||||:::||:|| ||
Db   1 MIEVLCNDRLGKKKVRVKCMPDDTVGDFKKLVAAQTGTDPRRIVLKKWHSVFKDNITLADY 60

Qy   61 EIHDGMNLELYY 72
         |||||:|::||:|
Db   61 EIHDGMSLEMYY 72

RESULT 5
ID   090461   PRELIMINARY;   PRT;   115 AA.
AC   090461;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   GP120 (FRAGMENT).
GN   ENV.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=THSC62;
RA   Subbarao S., Vanichseni S., Kitayaporn D., Hart C., Raktham S.,
RA   Young N., Srisuwanalai L., Luo C.C., Mastro T.D.;
RT   "Genetic Characterization of Incident HIV-1 Subtype B and E Strains
RT   from Injecting Drug Users (IDUs) in Bangkok, Thailand.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF081754; AAC34012.1; -.
DR   INTERPRO; IPR000777; -.
DR   PFAM; PF00516; GP120; 1.
FT   NON_TER   1   1
FT   NON_TER   115   115
SQ   SEQUENCE   115 AA;   13117 MW;   2E15C4A6A3405CD1 CRC64;

Query Match   22.6%; Score 89; DB 14; Length 115;
Best Local Similarity   29.5%; Pred. No. 0.0039;
Matches   28; Conservative   9; Mismatches 26; Indels 32; Gaps   4;

Qy   10 LGKKVRVKC-------------------NTDDTIGDLKKLIAAQTGTRWNKIV------ 43
          |||:||:|| |||:|| ||||::::||| ::
Db   13 LNKSVEINCTRPSNNTRTSIRIGPGQVFYRTGDIIGDIRKAYCEINGTKWNVLKQVTEK 72

Qy   44 LKKWY---TIFKDHVSLGDYEI---HDGMNLELYY 72
          ||| :   :|| ||||| |:| :|
Db   73 LKKHFNKTIVFEPPSGGDLEITTHHFNCRGEFFY 107

RESULT 6
Q9WNQ6
ID   Q9WNQ6   PRELIMINARY;   PRT;   114 AA.
AC   Q9WNQ6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN   ENV.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=THSC118;
RA   Subbarao S., Vanichseni S., Hu D.;
RT   "Genetic Characterization of Incident HIV-1 Subtype E and B Strains
RT   from a Prospective Cohort of Injecting Drug Users in Bangkok,
RT   Thailand.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF151752; AAD44294.1; -.
DR   INTERPRO; IPR000777; -.
DR   PFAM; PF00516; GP120; 1.
KW   Envelope protein.
FT   NON_TER   1   1
FT   NON_TER   114   114
SQ   SEQUENCE   114 AA;   13007 MW;   23C765F037B7CC5E CRC64;

Query Match   22.5%; Score 88.5; DB 14; Length 114;
Best Local Similarity   32.2%; Pred. No. 0.0044;
Matches   29; Conservative 12; Mismatches 30; Indels 19; Gaps   5;

Qy   2 IEVVC-----NDRLGKKV----RVKCNTDDTIGDLKKLIAAQTGTRWNKIV-----LKKW 47
         :|::| ||||||| ||: :| |||:|| |||| :||||||||:
Db   17 VEINCTRPSNNTRTGVTIGPGQVFYRTGEIIGDIRKAYCEVNGTKWNKVLKQVTEKLKEH 76

Qy   48 Y--TIFKDHVSLGDYEI---HDGMNLELYY 72
         : ||| |||| |:| :|
Db   77 YNKTIIFQPPSGGDLEITMHHFNCRGEFFY 106

RESULT 7
Q9JBC9
ID   Q9JBC9   PRELIMINARY;   PRT;   203 AA.
AC   Q9JBC9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN   ENV.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20027278; PubMed=10559367;
RA   Shankarappa R., Margolick J.B., Gange S.J., Rodrigo A.G., Upchurch D.,
RA   Farzadegan H., Gupta P., Rinaldo C.R., Learn G.H., He X., Huang X.-L.,
RA   Mullins J.I.;
RT   "Consistent viral evolutionary changes associated with the progression
RT   of human immunodeficiency virus type 1 infection.";
RL   J. Virol. 73:10489-10502(1999).
DR   EMBL; AF138608; AAF32933.1; -.
FT   NON_TER   1   1
FT   NON_TER   203   203
SQ   SEQUENCE   203 AA;   22963 MW;   A2BE13A5B8OC5CBA CRC64;

Query Match   22.3%; Score 88; DB 14; Length 203;
Best Local Similarity   37.1%; Pred. No. 0.01;
Matches   26; Conservative   7; Mismatches 23; Indels 14; Gaps   3;

Qy   7 NDRLGKKVRVKC------NTDDTIGDLKKLIAAQTGTRWNKI-----VLKKW------YTIFK 52
         ||:||: :|:|| :|||||||||||||||||||||| :|||| :||||
Db   34 NRHTGKSIRMGLGRAWHTTREIIGDIRKAYCTLNGTWNKILEQVVKKLREQFRNKTIVF 93

```
Qy        53  DHVSLGDYEI 62
              :| |  || ||
Db        94  NHSSGGDLEI 103


RESULT  8
Q71412
ID   Q71412                   PRELIMINARY;           PRT;     115 AA.
AC   Q71412;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN, V3 REGION (FRAGMENT).
GN   ENV.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TH100;
RX   MEDLINE=96014957; PubMed=7576318;
RA   Kalish M.L., Baldwin A., Raktham S., Wasi C., Luo C., Schochetman G.,
RA   Mastro T.D., Young N., Vanichseni S., Rubsamen-Waigmann H.,
RA   Von Briesen H., Mullins J.I., Delwart E., Herring B., Esparza J.,
RA   Heyward W.L., Osmanov S.;
RT   "The evolving molecular epidemiology of HIV-1 envelope subtypes in
RT   injecting drug users in Bangkok, Thailand: implications for HIV
RT   vaccine trials.";
RL   AIDS 9:851-857(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TH100;
RA   Siepel A.C.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U22542; AAA88941.1; -.
DR   INTERPRO; IPR000777; -.
DR   PFAM; PF00516; GP120; 1.
KW   Envelope protein.
FT   NON_TER       1      1
FT   NON_TER     115    115
SQ   SEQUENCE   115 AA;   13082 MW;  505DB9387E8A1918 CRC64;

Query Match             22.1%;  Score 87;  DB 14;  Length 115;
Best Local Similarity   29.2%;  Pred. No. 0.0067;
Matches    28;  Conservative    7;  Mismatches   27;  Indels   34;  Gaps    4;

Qy        10  LGKKVRVKC------------------NTDDTIGDLKKLIAAQTGTRWNKIVLKKWYT 49
              |:|  |||                  || ||:|||      :| ||| ||:|| ||
Db        13  LNKSVEINCTRPSSNTRTSIRIGPGQVFYKTGDIIGDIRKAYCEINGTKWNK-ALKQVAE 71

Qy        50  IFKDHV----------SLGDYEI---HDGMNLELY 72
              |:|              | ||| |   | |  |:|
Db        72  KFKEHFINNTITFQPPSGGDLEITTHHFNCRGEFFY 107


RESULT  9
O89553
ID   O89553                   PRELIMINARY;           PRT;     216 AA.
```

```
AC   O89553;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN   ENV.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=L6132;
RA   Chen Y.-M., Lee C.-M.;
RT   "Molecular epidemiology of HIV-1 infection in Taiwan.";
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF041121; AAC25046.1; -.
DR   INTERPRO; IPR000777; -.
DR   PFAM; PF00516; GP120; 1.
KW   Envelope protein.
FT   NON_TER       1      1
FT   NON_TER     216    216
SQ   SEQUENCE   216 AA;   24532 MW;  C5DDF30E6A65A6C7 CRC64;

Query Match             22.1%;  Score 87;  DB 14;  Length 216;
Best Local Similarity   33.8%;  Pred. No. 0.014;
Matches    26;  Conservative   12;  Mismatches   23;  Indels   16;  Gaps    4;

Qy         2  IEVVC---NDRLGKKVR-----VKCNTDDTIGDLKKLIAAQTGTRWNKIV------LKKW 47
              :| |     |  |:  :   :| : | |||||      :| ||| |||:|     ||||
Db        13  VEINCTRPSDNVRTSIRIGPGQVFYKTGDIIGDIRKAYCEINGTKWNKVLTQVTEKLKEH 72

Qy        48  Y--TIFKDHVSLGDYEI 62
              :  || |    | ||||
Db        73  FNKTIFFQPPSGGDLEI 89


RESULT  10
Q9WNN3
ID   Q9WNN3                   PRELIMINARY;           PRT;     115 AA.
AC   Q9WNN3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN   ENV.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=THSC78;
RA   Subbarao S., Vanichseni S., Hu D.;
RT   "Genetic Characterization of Incident HIV-1 Subtype E and B Strains
RT   from a Prospective Cohort of Injecting Drug Users in Bangkok,
RT   Thailand.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF151775; AAD44317.1; -.
DR   INTERPRO; IPR000777; -.
DR   PFAM; PF00516; GP120; 1.
KW   Envelope protein.
FT   NON_TER       1      1
FT   NON_TER     115    115
SQ   SEQUENCE   115 AA;   13144 MW;  F6A683AFFC5A467B CRC64;

Query Match             21.8%;  Score 86;  DB 14;  Length 115;
Best Local Similarity   28.9%;  Pred. No. 0.0088;
Matches    28;  Conservative    8;  Mismatches   27;  Indels   34;  Gaps    4;

Qy         9  RLGKKVRVKC------------------NTDDTIGDLKKLIAAQTGTRWNKIVLKKWY 48
              :|                         || ||:|||      :| ||| ||:|| |
Db        12  QLNKSVEINCTRPSNNTRTSITIGPGQVFYRTGDIIGDIRKAYCEINGTKWNK-VLKQVT 70
```

```
Qy       49 TIFKDHV---------SLGDYEI---HDGMNLELYY 72
               |:|           |||  ||  |   |:|
Db       71 EKLKEHFNNKTIIFQPPSGGDIEITMHHFNCKGEFFY 107

RESULT 11
ID   041876           PRELIMINARY;       PRT;    164 AA.
AC   041876;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN   ENV.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VN23;
RX   MEDLINE=96344000; PubMed=8738437;
RA   Nerurkar V.R., Nguyen H.T., Dashwood W.M., Hoffmann P.R., Yin C.,
RA   Morens D.M., Kaplan A.H., Detels R., Yanagihara R.;
RT   "HIV type 1 subtype E in commercial sex workers and injection drug
RT   users in southern Vietnam.";
RL   AIDS Res. Hum. Retroviruses 12:841-843(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VN23;
RA   Nerurkar V.R., Nguyen H.T., Woodward C.L., Hoffmann P.R.,
RA   Dashwood W.-M., Long H.T., Morens D.M., Detels R., Yanagihara R.;
RL   Cell. Mol. Biol. 0:0-0(1997).
DR   EMBL; U90087; AAB64376.1; -.
DR   INTERPRO; IPR000777; -.
DR   PFAM; PF00516; GP120; 1.
KW   Envelope protein.
FT   NON_TER        1      1
FT   NON_TER      164    164
SQ   SEQUENCE   164 AA;  18501 MW;  1D26498B47DA17C7 CRC64;

Query Match             21.8%;  Score 86;  DB 14;  Length 164;
  Best Local Similarity   28.4%;  Pred. No. 0.013;
  Matches   27;  Conservative   10;  Mismatches   26;  Indels   32;  Gaps    4;

Qy       10 LGKKVRVKC----------------NTDDTIGDLKKLIAAQTGTRWNKIV------- 43
               |  |                  ||  |||||||||||::|:::
Db       27 LNKSVEINCTRPSNNTRTSMTIGPGQVFYRTGDIIGDIRKAYCEINGTKWNEVLKQVAGK 86

Qy       44 LKKWY---TIFKDHVSLGDYEI---HDGMNLELYY 72
               ||:|     |||:||   |:|
Db       87 LKEHFNNKTIIFQPPSGGDLEITTHHFNCRGEFFY 121

RESULT 12
ID   Q71495           PRELIMINARY;       PRT;    114 AA.
AC   Q71495;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN, V3 REGION (FRAGMENT).
GN   ENV.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TH98;
RX   MEDLINE=96014957; PubMed=7576318;
RA   Kalish M.L., Baldwin A., Raktham S., Wasi C., Luo C., Schochetman G.,
RA   Mastro T.D., Young N., Vanichseni S., Rubsamen-Waigmann H.,
RA   Von Briesen H., Mullins J.I., Delwart E., Herring B., Esparza J.,
RA   Heyward W.L., Osmanov S.;
RT   "The evolving molecular epidemiology of HIV-1 envelope subtypes in
RT   injecting drug users in Bangkok, Thailand: implications for HIV
RT   vaccine trials.";
RL   AIDS 9:851-857(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TH98;
RX   MEDLINE=96014956; PubMed=7576317;
RA   Wasi C., Herring B., Raktham S., Vanichseni S., Mastro T.D.,
RA   Young N.L., Rubsamen-Waigmann H., Von Briesen H., Kalish M.L., Luo C.,
RA   Pau C., Baldwin A., Mullins J.I., Delwart E.L., Esparza J.,
RA   Heyward W.L. Osmanov S.;
RT   "Determination of HIV-1 subtypes in injecting drug users in Bangkok,
RT   Thailand, using peptide-binding enzyme immunoassay and heteroduplex
RT   mobility assay: evidence of increasing infection with HIV-1 subtype
RT   E.";
RL   AIDS 9:843-849(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TH98;
RA   Siepel A.C.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U22625; AAA89024.1; -.
DR   INTERPRO; IPR000777; -.
DR   PFAM; PF00516; GP120; 1.
KW   Envelope protein.
FT   NON_TER        1      1
FT   NON_TER      114    114
SQ   SEQUENCE   114 AA;  13191 MW;  3801EA986E40539E CRC64;

Query Match             21.7%;  Score 85.5;  DB 14;  Length 114;
  Best Local Similarity   37.1%;  Pred. No. 0.0099;
  Matches   26;  Conservative    6;  Mismatches   25;  Indels   13;  Gaps    3;

Qy       15 RVKCNTDDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKDHV--------SLGDYEI--- 62
            || :||| |||||||||| |:||| ||| ||:|  |:|          |:|
Db       38 RVYYKTGDIIGDIRKAYCEINGTKWNK-VLKQVTEKLKEHFNRTILFQPPSGGDLEITMH 96

Qy       63 HDGMNLELYY 72
              |:| |:||
Db       97 HFNCRGEFFY 106

RESULT 13
ID   090467           PRELIMINARY;       PRT;    114 AA.
AC   090467;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   GP120 (FRAGMENT).
GN   ENV.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=THSC70;
RA   Subbarao S., Vanichseni S., Kitayaporn D., Hart C., Raktham S.,
RA   Young N., Srisuwanalai L., Luo C.C., Mastro T.D.;
RT   "Genetic Characterization of Incident HIV-1 Subtype B and E Strains
RT   from Injecting Drug Users (IDUs) in Bangkok, Thailand.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF081761; AAC34019.1; -.
DR   INTERPRO; IPR000777; -.
DR   PFAM; PF00516; GP120; 1.
FT   NON_TER        1      1
FT   NON_TER      114    114
SQ   SEQUENCE   114 AA;  12973 MW;  F4CF55404ED58FC4 CRC64;
```

```
Query Match             21.7%;  Score 85.5;  DB 14;  Length 114;
Best Local Similarity   29.5%;  Pred. No. 0.0099;
Matches   28;  Conservative    7;  Mismatches    27;  Indels    33;  Gaps    4;

Qy    10 LGKKVRVKC------------NTDDTIGDLKKLIAAQTGTRWNKIVLKKWYT 49
          | | |:| |            | |!!:!|    ||:||| |||:
Db    13 LNKSVEINCTRPSNNTRTSVPIGPGQVFYRTGDIIGDIRKAYCEINGTKWNK-VLKQVAE 71

Qy    50 IFKDHV-------SLGDYEI---HDGMNLELYY 72
          :|:|          | |||     | :|
Db    72 KLKEHFNKTIIFQPPSGGDLEITMHHFNCKGEFFY 106

RESULT 14
ID   Q9QCN8        PRELIMINARY;       PRT;    176 AA.
AC   Q9QCN8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   ENVELOPE GLYCOPROTEIN (FRAGMENT).
DE   ENV.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KR523;
RA   Lee J.S., Nam J.G., Kim E.Y., Kang C., Koo B.K., Cho H.W.;
RT   "HIV-1 subtype E virus in South Korea.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF204263; AAF13288.1; -.
DR   INTERPRO; IPR000777; -.
DR   PFAM; PF00516; GP120; 1.
KW   Envelope protein.
FT   NON_TER       1      1
FT   NON_TER     176    176
SQ   SEQUENCE   176 AA;  19562 MW;  54ABFCBF7727357 CRC64;


Query Match             21.7%;  Score 85.5;  DB 14;  Length 176;
Best Local Similarity   29.5%;  Pred. No. 0.017;
Matches   28;  Conservative    7;  Mismatches    27;  Indels    33;  Gaps    4;

Qy    10 LGKKVRVKC------------NTDDTIGDLKKLIAAQTGTRWNKIVLKKWYT 49
          | | |:| |            | |!!:!|    ||:||| |||:
Db    15 LNKSVEINCTRPSNNTRTSIPIGPGQVFYKTGDIIGDIRKAYCEINGTKWNK-VLKQVTE 73

Qy    50 IFKDHV-------SLGDYEI---HDGMNLELY 72
          :|:|          | |||     | :|
Db    74 KLKEHFNKTISFQPPSGGDLEITMHHFNCGGEFFY 108

RESULT 15
ID   Q75118        PRELIMINARY;       PRT;    209 AA.
AC   Q75118;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ENVELOPE, C2-V5 REGION (FRAGMENT).
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CMU08;
RA   McCutchan F.E., Artenstein A.W., Sanders-Buell E., Slaminen M.O.,
RA   Carr J.K., Mascola J.R., Yu X.F., Nelson K.E., Khamboonruang C.,
RA   Schmitt D., Kieny M.P., McNeil J.G., Burke D.S.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U48268; AAB06207.1; -.
```

```
DR   INTERPRO; IPR000777; -.
DR   PFAM; PF00516; GP120; 1.
FT   NON_TER       1      1
FT   NON_TER     209    209
SQ   SEQUENCE   209 AA;  23168 MW;  8FC649F7D30B3469 CRC64;


Query Match             21.7%;  Score 85.5;  DB 14;  Length 209;
Best Local Similarity   31.9%;  Pred. No. 0.02;
Matches   29;  Conservative   10;  Mismatches    31;  Indels    21;  Gaps    4;

Qy     2 IEVCNDRLGKK------------VRVKCNTDTIGDLKKLIAAQTGTRWNKIVLKKWYTIFKD 53
          :|:| |       | :| | :|||:||      ||:||||
Db    40 VEINCTRPSNKKRQRVTIGPGRVYYSTGDIMGDIRKAHCEINGTKWNK-VLKQVTEKLKE 98

Qy    54 HV---------SLGDYEI---HDGMNLELYY 72
          | |         | ||| |   | :|
Db    99 HFNKTIVFQPPSGGDLEITMHHFNCRGEFFY 129


Search completed: June 13, 2001, 20:29:04
Job time: 211 sec
```

```
                    GenCore version 4.5
            Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      June 13, 2001, 20:24:28 ; Search time 34.18 Seconds
                                        (without alignments)
                                        73.161 Million cell updates/sec

Title:       US-09-331-930A-2
Perfect score: 394
Sequence:    1 MIEVVCNDRLGKKVRVKCNT.........HVSLGDYEIHDGMNLELYYQ 73

Scoring table: BLOSUM62
             Gapop 10.0 , Gapext 0.5

Searched:    93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters:      93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SwissProt_39:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                           SUMMARIES

                  %
Result         Query
No.    Score   Match  Length  DB  ID            Description
-----  -----  -----  ------  --  -----------  ------------------
1      73.5   18.7   949     1   HYSA_STRPN   Q54873 streptococc
2      65.5   16.6   504     1   C7B7_ARATH   Q96514 arabidopsis
3      65.5   16.6   797     1   CTPE_MYCTU   O08365 mycobacteri
4      65     16.5   333     1   Y642_METJA   Q58059 methanococc
5      65     15.7   527     1   RAG2_RABIT   P34089 oryctolagus
6      62     15.7   662     1   ABFA_STRLI   P53627 streptomyce
7      61     15.5   177     1   VC08_VACCV   P17364 vaccinia vi
8      60.5   15.4   338     1   ANX2_CHICK   P17785 gallus gall
9      60.5   15.4   1275    1   GNRP_HUMAN   Q13972 homo sapien
10     60     15.2   69      1   TRY1_SALTI   P12059 salmonella
11     60     15.2   740     1   PSAB_SYNEN   P25897 synechococc
12     59.5   15.1   644     1   YHOB_DROME   P12258 drosophila
13     59     15.0   669     1   ENV_MLVAV    P03386 akv murine
14     59     15.0   730     1   PSAB_SYNY3   P29255 synechocyst
15     59     15.0   732     1   PSAB_SYNP2   P17155 synechococc
16     58.5   14.8   330     1   CATS_RAT     Q02765 rattus norv
17     58.5   14.8   861     1   ENV_HV1KB    P31819 human immun
18     58     14.7   93      1   UBIL_NPVOP   Q05120 orgyia pseu
19     58     14.7   184     1   VC08_VACCC   P21041 vaccinia vi
20     58     14.7   381     1   Y438_METJA   Q57880 methanococc
21     58     14.7   734     1   PSAB_PORPU   P51285 porphyra pu
22     57.5   14.6   185     1   PTH_RICPR    Q92cv4 rickettsia
23     57.5   14.6   339     1   ANX2_XENLA   P24801 xenopus lae
24     57.5   14.6   343     1   HRCA_MYCGE   P47447 mycoplasma
25     57.5   14.6   661     1   MX1_HUMAN    P20591 homo sapien
26     57.5   14.6   732     1   ERG7_HUMAN   P48449 homo sapien
27     57.5   14.6   1024    1   SYIP_STAAU   P41368 staphylococ
28     57     14.5   267     1   CBBQ_PSEHY   Q51858 pseudomonas
29     57     14.5   434     1   TBBQ_YEAST   Q99867 homo sapien
30     56.5   14.3   222     1   ERG2_YEAST   P32352 saccharomyc
31     56.5   14.3   338     1   ANX2_MOUSE   P07356 mus musculu
32     56.5   14.3   468     1   SYFA_SULSO   P95961 sulfolobus
33     56     14.2   414     1   K1M8_HUMAN   O76015 homo sapien
```

```
34   14.2   56     496  1   MGLA_TREPA   Q56342 treponema p
35   14.2   56     734  1   PSAB_GUITH   O07507 guillardia
36   14.2   56     1262 1   GNRP_MOUSE   P27671 mus musculu
37   14.2   56     4486 1   DYH9_HUMAN   Q9nyc9 homo sapien
38   14.1   55.5   261  1   RS3A_ORYSA   P49397 oryza sativ
39   14.1   55.5   277  1   VA38_VARV    P33853 variola vir
40   14.1   55.5   282  1   NADA_APLCA   P29241 aplysia cal
41   14.1   55.5   290  1   YEAB_BACSU   P46348 bacillus su
42   14.1   55.5   867  1   SFMD_ECOLI   P77468 escherichia
43   14.1   55.5   933  1   MSH2_RAT     P54275 rattus norv
44   14.1   55.5   1008 1   VGLM_UUK     P09613 uukuniemi v
45   14.0   55     267  1   LPXA_PROMI   P72215 proteus mir
```

                           ALIGNMENTS

```
RESULT  1
HYSA_STRPN
ID   HYSA_STRPN     STANDARD;      PRT;     949 AA.
AC   Q54873; Q54874;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYALURONATE LYASE PRECURSOR (EC 4.2.2.1) (HYALURONIDASE) (HYASE).
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYPE 23;
RX   MEDLINE=94156460; PubMed=8112843;
RA   Berry A.M., Lock R.A., Thomas S.M., Rajan D.P., Hansman D.,
RA   Paton J.C.;
RT   "Cloning and nucleotide sequence of the Streptococcus pneumoniae
RT   hyaluronidase gene and purification of the enzyme from recombinant
RL   Escherichia coli.";
RL   Infect. Immun. 62:1101-1108(1994).
RN   [2]
RP   CRYSTALLIZATION.
RX   MEDLINE=98234706; PubMed=9573623;
RA   Jedrzejas M.J., Chantalat L., Mewbourne R.B.;
RT   "Crystallization and preliminary X-ray analysis of Streptococcus
RT   pneumoniae hyaluronate lyase.";
RL   J. Struct. Biol. 121:73-75(1998).
CC   -!- CATALYTIC ACTIVITY: HYALURONATE = N 3-(4-DEOXY-BETA-D-GLUC-4-
CC       ENURONOSYL)-N-ACETYL-D-GLUCOSAMINE.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC   -!- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 8.
CC   -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC       IN THE REGION OF THE MEMBRANE ANCHOR.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; L20670; AAA53685.1; -.
DR   EMBL; L20670; AAA53686.1; ALT_INIT.
DR   InterPro; IPR001899; -.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW   Cell wall; Transmembrane; Lyase; Signal.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?     949      HYALURONATE LYASE.
FT   DOMAIN        1     924      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    925     943      MEMBRANE ANCHOR (POTENTIAL).
FT   DOMAIN      944     949      CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   949 AA;  107756 MW;  77FE0AC7A765C3E2 CRC64;
```

Query Match   18.7%; Score 73.5; DB 1; Length 949;
Best Local Similarity   32.8%; Pred. No. 0.79;
Matches 22; Conservative 9; Mismatches 27; Indels 9; Gaps 3;

```
Qy   13 KVRVKCNTDDTIGDLKKLIAAQTG--TRWNKIV--LKKWYTIFKDHVSLGDYEIHDGM 66
         |:|:||:||:::|  ::|    ||  ||:|||  ||::|:|:||||||:||||:
Db   11 KLRFKIKTDNKVGIAKVRIIEESGKDKRLWNSATTSGTKDWQTIEADYSPTLDV---DKI 67

Qy   67 NLELYQ 73
         |||:|:
Db   68 KLELFYE 74
```

RESULT 2
C7B7_ARATH
ID   C7B7_ARATH   STANDARD;   PRT;   504 AA.
AC   Q96514;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   CYTOCHROME P450 71B7 (EC 1.14.-.-).
GN   CYP71B7 OR F3F19.13.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=97288381; PubMed=9143359;
RA   Maughan J.A., Nugent J.H.A., Hallahan D.L.;
RT   "Expression of CYP71B7, a cytochrome P450 expressed sequence Tag from
RT   Arabidopsis thaliana.";
RL   Arch. Biochem. Biophys. 341:104-111(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lee J.M., Lenz C.,
RA   Liu S., Li I., Kremenetskaia I., Luros J., Ngan I., Gonzalez A.,
RA   Altafi H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,
RA   Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA   Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ROSETTE LEAVES. ALSO
CC   EXPRESSED IN ROOTS, LEAVES, FLOWERS AND SILIQUES.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X97864; CAA66458.1; -.
DR   EMBL; AC007357; AAD31064.1; -.
DR   InterPro; IPR001128; -.
DR   InterPro; IPR002401; -.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00463; EP450I.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Heme.
FT   BINDING   446   446   HEME (BY SIMILARITY).
SQ   SEQUENCE   504 AA; 57209 MW; B310296ICF954FBA CRC64;

Query Match   16.6%; Score 65.5; DB 1; Length 504;
Best Local Similarity   28.0%; Pred. No. 3.4;
Matches 21; Conservative 16; Mismatches 16; Indels 23; Gaps 3;

```
Qy    1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKKLIAAQ--TGTRWNKIVLKKWYTIF------ 51
        | |:||:  :|: ||::|| |:|:|||||:|:|:    :||:|||:||||||| :
Db  319 MAELIRNPRVMKKVQDEIRT--TLGDKKQRITEQDLSQVHYFKLVVKEIFRLHPAAPLLL 376

Qy   52 ---KDHVSLGDYEI 62
           ::|:|:||||:|
Db  377 PRETMSHVKIQGYDI 391
```

RESULT 3
CTPE_MYCTU
ID   CTPE_MYCTU   STANDARD;   PRT;   797 AA.
AC   O08365;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PROBABLE CATION-TRANSPORTING ATPASE E (EC 3.6.1.-).
GN   CTPE OR RV0908 OR MTCY21C12.02.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Sqares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC   (E1-E2 ATPASES).
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; Z95210; CAB08506.1; -.
DR   TubercuList; Rv0908; -.
DR   InterPro; IPR001757; -.
DR   Pfam; PF00122; E1-E2_ATPase; 2.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
KW   Hydrolase; Transmembrane; Phosphorylation; ATP-binding.
FT   TRANSMEM   55   75   POTENTIAL.
FT   TRANSMEM   215   235   POTENTIAL.
FT   TRANSMEM   254   274   POTENTIAL.
FT   TRANSMEM   549   569   POTENTIAL.
FT   TRANSMEM   601   621   POTENTIAL.
FT   TRANSMEM   633   653   POTENTIAL.
FT   TRANSMEM   667   687   POTENTIAL.
FT   TRANSMEM   703   723   POTENTIAL.
FT   TRANSMEM   729   749   POTENTIAL.
FT   TRANSMEM   764   784   POTENTIAL.
FT   MOD_RES   301   301   PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   797 AA; 84973 MW; 4C5034FC6052FC7B CRC64;

Query Match   16.6%; Score 65.5; DB 1; Length 797;
Best Local Similarity   29.6%; Pred. No. 5.6;
Matches 24; Conservative 14; Mismatches 22; Indels 21; Gaps 5;

```
Qy    4 VVCNDRLG--KKVRVKCNTDD--TIGDKLKLIAAQTGTR--WNKIVLKKWYTIFKDHVS 56
        ::  :   :::||| | :||   ||  ||| |  |:|||   |:::||
Db   74 IIANSVIGMVQEIRAKQTLDKLAIIGQAKPLVRRQSGTRTRSTNEVVL------DDIIE 126

Qy   57 LG------DYEIHDGMNLEL 70
        ||       ||: |||
Db  127 LGPGDQVVVDGEVVEEENLEI 147

RESULT 4
Y642_METJA
ID   Y642_METJA      STANDARD;      PRT;      333 AA.
AC   Q58059;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ0642.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SIMILARITY: TO THE C-TERMINAL OF M.JANNASCHII MJ0872 AND MJ0878.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; U67512; AAB98644.1; -.
DR   TIGR; MJ0642; -.
DR   InterPro; IPR002491; -.
DR   Pfam; PF01497; Peripla_BP_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   333 AA;   40005 MW;   1EF47CDE74593BDF CRC64;

Query Match      16.5%;   Score 65;   DB 1;   Length 333;
Best Local Similarity   37.9%;   Pred. No. 2.4;
Matches   11; Conservative   8; Mismatches   10; Indels 0; Gaps 0;

Qy   27 LKKLIAAQTGTRWNKIVLKKWYTIFKDHV 55
        ||:|:|  |  |||||:|| ::|:::
Db   62 IKKMIKSIKGVFWADFVLKKYYPLYSDYL 90

RESULT 5
RAG2_RABIT
ID   RAG2_RABIT      STANDARD;      PRT;      527 AA.
AC   P34089;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   V(D)J RECOMBINATION ACTIVATING PROTEIN 2 (RAG-2).
GN   RAG2.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93354283; PubMed=8350872;
RA   Fuschiotti P., Harindranath N., Mage R.G., McCormack W.T.,
RA   Dhanarajan P., Roux K.H.;
RT   "Recombination activating genes-1 and -2 of the rabbit: cloning and
RT   characterization of germline and expressed genes.";
RL   Mol. Immunol. 30:1021-1032(1993).
CC   -!- FUNCTION: DURING LYMPHOCYTE DEVELOPMENT, THE GENES ENCODING
CC       IMMUNOGLOBULINS AND T CELL RECEPTORS ARE ASSEMBLED FROM VARIABLE
CC       (V), DIVERSITY (D), AND JOINING (J) GENE SEGMENTS. THIS
CC       COMBINATORIAL PROCESS, KNOWN AS V(D)J RECOMBINATION, ALLOWS THE
CC       GENERATION OF AN ENORMOUS RANGE OF BINDING SPECIFICITIES FROM A
CC       LIMITED AMOUNT OF GENETIC INFORMATION. THE RAG1/RAG2 COMPLEX
CC       INITIATES THIS PROCESS BY BINDING TO THE CONSERVED RECOMBINATION
CC       SIGNAL SEQUENCES (RSS) AND INTRODUCING A DOUBLE-STRAND BREAK
CC       BETWEEN THE RSS AND THE ADJACENT CODING SEGMENT. THESE BREAKS ARE
CC       GENERATED IN TWO STEPS, NICKING OF ONE STRAND (HYDROLYSIS),
CC       FOLLOWED BY HAIRPIN FORMATION (TRANSESTERIFICATION). RAG1/2 HAS
CC       ALSO BEEN SHOWN TO FUNCTION AS A TRANSPOSASE IN VITRO, AND TO
CC       POSSESS RSS-INDEPENDENT ENDONUCLEASE ACTIVITY (END PROCESSING) AND
CC       HAIRPIN OPENING. RAG1 ALONE CAN BIND TO RSS BUT STABLE, EFFICIENT
CC       BINDING REQUIRES RAG2. ALL KNOWN CATALYTIC ACTIVITIES REQUIRE THE
CC       PRESENCE OF BOTH PROTEINS (BY SIMILARITY).
CC   -!- SUBUNIT: THE RAG COMPLEXES APPEAR TO CONTAIN THREE TO FIVE
CC       MOLECULES OF RAG2 FOR EACH MOLECULE OF RAG1.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: THYMUS.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; M77667; AAA03029.1; -.
DR   PIR; S42512; S42512.
KW   Hydrolase; Endonuclease; Nuclear protein; DNA-binding;
KW   DNA recombination.
FT   DOMAIN       352    412       ASP/GLU-RICH (ACIDIC).
SQ   SEQUENCE   527 AA;   59491 MW;   B8BDC0BD709AE7ED CRC64;

Query Match      15.7%;   Score 62;   DB 1;   Length 527;
Best Local Similarity   20.6%;   Pred. No. 9;
Matches   21; Conservative 18; Mismatches   29; Indels 34; Gaps 4;

Qy    1 MIEVVCNDRLGKKVRVKCNTDDTIGDLKK-----------LIAAQTGTRWNK 41
        :: ||:  ||| | ||| | |||||::|               :: :||
Db  109 VMSVVCKN--NKKVTFRCTERDLVGDVPEARYGHSLDVVYSRGKSMGVLFGGRSYMPSNQ 166

Qy   42 IVLKKWYTIFK-DHVSLGDY----------EIHDGMNLEL 70
         ||::  ::|   |:: | |          :|||||:||
Db  167 RTTEKWNSVADCLPHVFLVDFEFGCATSYILPELQDGLSFHV 208

RESULT 6
ABFA_STRLI
ID   ABFA_STRLI      STANDARD;      PRT;      662 AA.
AC   P53627;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ALPHA-L-ARABINOFURANOSIDASE (EC 3.2.1.55) (ARABINOSIDASE) (ABF)
DE   (ALPHA-L-AF).
GN   ABFA.
```

```
OS   Streptomyces lividans.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1916;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-23.
RC   STRAIN=66 / 1326;
RX   MEDLINE=94379974; PubMed=8092996;
RA   Manin C., Shareek F., Morosoli R., Kluepfel D.;
RT   "Purification and characterization of an alpha-L-arabinofuranosidase
RT   from Streptomyces lividans 66 and DNA sequence of the gene (abfA).";
RL   Biochem. J. 302:443-449(1994).
CC   -!- FUNCTION: IMPORTANT ROLE IN THE DEGRADATION OF ARABINOSE-
CC       CONTAINING HEMICELLULOSES. HYDROLYSES RAPIDLY THE SHORT-CHAIN
CC       ARABINO-OLIGOXYLOSIDES FROM DIGESTION OF XYLAN WITH XYLANASES.
CC       ACTS SLOWLY ON ARABINAN AND ARABINOXYLAN FROM WHEAT AND RYE FLOUR
CC       TO RELEASE L-ARABINOFURANOSE. DOES NOT ACT ON OAT-SPELTS XYLAN OR
CC       ARABINOGALACTAN. OPTIMAL ENZYME ACTIVITY OCCURS AT 60 DEGREES
CC       CELSIUS AND PH 6.0.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING ALPHA-L-
CC       ARABINOFURANOSIDE RESIDUES IN ALPHA-L-ARABINOSIDES.
CC   -!- SUBUNIT: OLIGOMER.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO FAMILY 51 OF GLYCOSYL HYDROLASES.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U04630; AAA61708.1; -.
KW   Hydrolase; Glycosidase.
SQ   SEQUENCE   662 AA;  72496 MW;  DAAF66A577C1D6D1 CRC64;

Query Match             15.7%;  Score 62;  DB 1;  Length 662;
Best Local Similarity   45.5%;  Pred. No. 12;
Matches   15;  Conservative    4;  Mismatches    4;  Indels   12;  Gaps    2;

Qy     34 QTGTRWNKIVLKKWYTIFKDHVSI-GDYEIHDG 65
           :| |||||:|| |:||||||||-||:|| |||
Db    221 ETFAEWEATVLKETYDLV-DHISLHAYYEPHDG 252


RESULT   7
VC08_VACCV
ID   VC08_VACCV     STANDARD;       PRT;   177 AA.
AC   P17364;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   PROTEIN C8.
GN   C8L.
OS   Vaccinia virus (strain WR).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10254;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89073756; PubMed=2849238;
RA   Kotwal G.J., Moss B.;
RT   "Analysis of a large cluster of nonessential genes deleted from a
RT   vaccinia virus terminal transposition mutant.";
RL   Virology 167:524-537(1988).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M22812; AAA69600.1; -.
DR   PIR; I31829; WZVZA9.
KW   Early protein.
SQ   SEQUENCE   177 AA;  20753 MW;  8C536DD9F90F0065 CRC64;

Query Match             15.5%;  Score 61;  DB 1;  Length 177;
Best Local Similarity   44.8%;  Pred. No. 3.6;
Matches   13;  Conservative    4;  Mismatches   10;  Indels   10;  Gaps    1;

Qy     43 VLKKWYTIFKDHVSLGDYEIHDGMNLELY 71
           |:|| ||||||: :|||||||| |:::||
Db     67 VMTKWCAPFIDTVSV--YTSHDNLNIEFY 93


RESULT   8
ANX2_CHICK
ID   ANX2_CHICK     STANDARD;       PRT;   338 AA.
AC   P17785;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ANNEXIN II (LIPOCORTIN II) (CALPACTIN I HEAVY CHAIN) (CHROMOBINDIN 8)
DE   (P36) (PROTEIN I) (CALPACTIN I HEAVY CHAIN) (CHROMOBINDIN 8)
DE   (P36) (PROTEIN I) (PLACENTAL ANTICOAGULANT PROTEIN IV) (PAP-IV).
GN   ANXA2 OR ANX2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90332221; PubMed=2143014;
RA   Gerke V., Koch W.;
RT   "The cDNA sequence of chicken annexin II.";
RL   Nucleic Acids Res. 18:4246-4246(1990).
RN   [2]
RP   SEQUENCE OF 1-69.
RX   MEDLINE=88296872; PubMed=2456953;
RA   Johnsson N., Johnsson K., Weber K.;
RT   "A discontinuous epitope on p36, the major substrate of src tyrosine-
RT   protein-kinase, brings the phosphorylation site into the
RT   neighbourhood of a consensus sequence for Ca2+/lipid-binding
RT   proteins.";
RL   FEBS Lett. 236:201-204(1988).
RN   [3]
RP   SEQUENCE OF 1-29.
RX   MEDLINE=89052656; PubMed=2973411;
RA   Johnsson N., Marriott G., Weber K.;
RT   "p36, the major cytoplasmic substrate of src tyrosine protein kinase,
RT   binds to its p11 regulatory subunit via a short amino-terminal
RT   amphiphatic helix.";
RL   EMBO J. 7:2435-2442(1988).
CC   -!- FUNCTION: CALCIUM-REGULATED MEMBRANE-BINDING PROTEIN WHOSE
CC       AFFINITY FOR CALCIUM IS GREATLY ENHANCED BY ANIONIC PHOSPHOLIPIDS.
CC       IT BINDS TWO CALCIUM IONS WITH HIGH AFFINITY.
CC   -!- SUBUNIT: TETRAMER OF 2 LIGHT CHAINS (P10 PROTEINS) AND 2 HEAVY
CC       CHAINS (P36 PROTEINS).
CC   -!- SUBCELLULAR LOCATION: IN THE LAMINA BENEATH THE PLASMA MEMBRANE.
CC   -!- DOMAIN: CONTAINS FOUR HOMOLOGOUS REPEATS WITH A CONSENSUS
CC       SEQUENCE COMMON TO ALL ANNEXIN PROTEINS. A PAIR OF THESE REPEATS
CC       MAY FORM ONE BINDING SITE FOR CALCIUM AND PHOSPHOLIPID.
CC   -!- MISCELLANEOUS: IT MAY CROSS-LINK PLASMA MEMBRANE PHOSPHOLIPIDS
CC       WITH ACTIN AND THE CYTOSKELETON AND BE INVOLVED WITH EXOCYTOSIS.
CC   -!- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; X53334; CAA37421.1; -.
DR   PIR; S10501; LUCH2.
DR   HSSP; P04083; 1BO9.
DR   InterPro; IPR001464; -.
DR   InterPro; IPR002389; -.
DR   Pfam; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
DR   PRINTS; PR00198; ANNEXINII.
DR   PROSITE; PS00223; ANNEXIN; 4.
KW   Annexin; Calcium/phospholipid-binding; Repeat; Phosphorylation.
FT   INIT_MET      0      0
FT   DOMAIN        1     23       P10 BINDING SITE (POTENTIAL).
FT   REPEAT       41    101       ANNEXIN 1.
FT   REPEAT      113    173       ANNEXIN 2.
FT   REPEAT      198    258       ANNEXIN 3.
FT   REPEAT      273    333       ANNEXIN 4.
FT   MOD_RES      25     25       PHOSPHORYLATION (BY SRC).
FT   MOD_RES      25     25       PHOSPHORYLATION (BY PKC).
SQ   SEQUENCE    338 AA;  38509 MW;  4A869FE9B1983C30 CRC64;

Query Match                 15.48%;  Score  60.5;  DB 1;  Length 338;
Best Local Similarity       29.48%;  Pred. No. 8.3;
Matches   25;  Conservative    9;  Mismatches   24;  Indels   27;  Gaps   4;

Qy    1 MIEVVC---NDRLGKKVRV--------KCNTDDTIGDLKKL-IAAQTGTRWNKIVLKK 46
          :|::|| -| -|| |  ||  || ||  :| |   |
Db  127 LIEIICSRTNQELNEINRVYREMYKTELEKDIISDTSGDFRKLMVALAKGKRC------ 179

Qy   47 WYTIFKDHVSLGDYEIHDGMNLELY 71
          :  |: |||:| |||
Db  180 ------EDTSVIDYELIDQDARELY 198

RESULT 9
GNRP_HUMAN
ID   GNRP_HUMAN     STANDARD;      PRT;   1275 AA.
AC   Q13972;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   GUANINE NUCLEOTIDE RELEASING PROTEIN (GNRP) (RAS-SPECIFIC NUCLEOTIDE
DE   EXCHANGE FACTOR CDC25).
     RASGRF1 OR CDC25.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wei W., Broek D.;
RT   "Cloning and analysis of the full length human cdc25 cDNA, a ras-
RT   specific nucleotide exchange factor.";
RL   Submitted (DEC-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC   -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC   -!- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; L26584; AAA58417.1; -.
```

```
DR   InterPro; IPR000048; -.
DR   InterPro; IPR000219; -.
DR   InterPro; IPR000651; -.
DR   InterPro; IPR001131; -.
DR   InterPro; IPR001849; -.
DR   InterPro; IPR001895; -.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF00169; PH; 2.
DR   Pfam; PF00617; RasGEF; 1.
DR   Pfam; PF00618; RasGEFN; 1.
DR   Pfam; PF00621; RhoGEF; 1.
DR   PROSITE; PS00720; GDS_CDC25; 1.
DR   PROSITE; PS00741; GDS_CDC24; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 2.
DR   PROSITE; PS50096; IQ; 1.
KW   Guanine-nucleotide releasing factor.
FT   DOMAIN       22    129       PH 1.
FT   DOMAIN      204    229       IQ.
FT   DOMAIN      247    431       DH.
FT   DOMAIN      467    584       PH 2.
FT   DOMAIN     1038   1272       RASGEF.
SQ   SEQUENCE   1275 AA;  145381 MW;  86C6F54AA1E451F1 CRC64;

Query Match                 15.48%;  Score  60.5;  DB 1;  Length 1275;
Best Local Similarity       32.0%;   Pred. No. 36;
Matches   16;  Conservative   12;  Mismatches   17;  Indels    5;  Gaps   3;

Qy    7 NDRLGKKVRVKCNTDDTIGDLKKLIAAQTGTR-WNKIVLKKWYTIFKDHV 55
          :|: | :|: |||:| |:||| ::|| :|| |:
Db  185 NERIQSTQTVAPNDEDS--DIKKIKKVQSFLRGW--LCRRKWKTIIQDYI 230

RESULT 10
TRY1_SALTI
ID   TRY1_SALTI     STANDARD;      PRT;     69 AA.
AC   P12059;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   TRAY PROTEIN.
GN   TRAY.
OS   Salmonella typhi.
OG   Plasmid IncFV pED208.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87056998; PubMed=2877970;
RA   Finlay B.B., Frost L.S., Paranchych W.;
RT   "Nucleotide sequence of the tra YALE region from IncFV plasmid
RT   pED208.";
RL   J. Bacteriol. 168:990-998(1986).
CC   -!- FUNCTION: INVOLVED IN THE CONJUGATION PROCESS OF BACTERIAL CELLS
CC       FOR THE EXCHANGE OF PLASMID DNA. IT IS ALSO RESPONSIBLE FOR
CC       CONJUGAL DNA METABOLISM. TRAY IS REQUIRED FOR STRAND-SPECIFIC
CC       NICKING AT ORIT, THE TRANSFER ORIGIN.
CC   -!- SIMILARITY: 26%, 32% AND 25% SIMILARITY TO TRAY PROTEINS OF
CC       PLASMID F, R1 AND R100, RESPECTIVELY IN THE REGION OF POS. 10-60.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the European Bioinformatics Institute.
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; M14733; AAA25604.1; -.
DR   EMBL; M14733; AAA25605.1; ALT_TERM.
DR   EMBL; M14733; AAA25606.1; ALT_INIT.
DR   PIR; A25161; A25161.
```

```
KW   Plasmid; Conjugation; DNA-binding.
SQ   SEQUENCE   69 AA;  7788 MW;  C600BCCC94FD278E CRC64;

Query Match             15.2%;  Score 60;  DB 1;  Length 69;
Best Local Similarity   37.5%;  Pred. No. 1.7;
  Matches   12;  Conservative    8;  Mismatches   10;  Indels    2;  Gaps    1;

Qy        10 LGKKVRVKCNTDDTIGDLKKLIAAQTGTRWNK 41
                :||||:|:|:||:||:||:|.||  :| |:|
Db        10 IGKKVNISCSLDEAIDELLMESALKSG--WSK 39

RESULT 11
PSAB_SYNEN
ID   PSAB_SYNEN     STANDARD;      PRT;    740 AA.
AC   P25897;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   PHOTOSYSTEM I P700 CHLOROPHYLL A APOPROTEIN A2.
OS   Synechococcus elongatus naegeli, and Synechococcus vulcanus.
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=1141, 32053;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.e.naegeli;
RX   MEDLINE=93252282; PubMed=8486290;
RA   Muehlenhoff U., Haehnel W., Witt H.T., Herrmann R.G.;
RT   "Genes encoding eleven subunits of photosystem I from the
RT   thermophilic cyanobacterium Synechococcus sp.";
RL   Gene 127:71-78(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.vulcanus;
RX   MEDLINE=92216055; PubMed=1558952;
RA   Shimizu T., Hiyama T., Ikeuchi M., Inoue Y.;
RT   "Nucleotide sequences of the psaA and psaB genes encoding the
RT   photosystem I core proteins from the thermophilic cyanobacterium
RT   Synechococcus vulcanus.";
RL   Plant Mol. Biol. 18:785-791(1992).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (4.0 ANGSTROMS).
RC   SPECIES=S.e.naegeli;
RX   MEDLINE=97057537; PubMed=8901876;
RA   Krauss N., Schubert W.D., Klukas O., Fromme P., Witt H.T., Saenger W.;
RT   "Photosystem I at 4-A resolution represents the first structural
RT   model of a joint photosynthetic reaction centre and core antenna
RT   system.";
RL   Nat. Struct. Biol. 3:965-973(1996).
CC   -!- FUNCTION: PSAA ALONG WITH PSAB BIND P700, THE PRIMARY ELECTRON
CC       DONOR OF PHOTOSYSTEM I (PSI) AS WELL AS THE ELECTRON ACCEPTORS A0,
CC       A1, AND FX.
CC   -!- COFACTOR: P700 IS A CHLOROPHYLL DIMER, A0 IS CHLOROPHYLL, A1 IS A
CC       PHYLLOQUINONE AND FX IS A 4FE-4S IRON-SULFUR CENTER.
CC   -!- SUBUNIT: HETERODIMER OF A1 AND A2 SUBUNITS WHICH BIND TOGETHER THE
CC       IRON-SULFUR CENTER FX.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
```

```
DR   EMBL; X63768; CAA45305.1; -.
DR   EMBL; D10986; BAA01760.1; -.
DR   PIR; S20923; S20923.
DR   PDB; 2PPS; 27-MAY-98.
DR   InterPro; IPR001280; -.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTOSYPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAB; 1.
KW   Chlorophyll; Photosystem I; Electron transport; Photosynthesis;
KW   Transmembrane; Iron-sulfur; 4Fe-4S; 3D-structure.
FT   INIT_MET        0      0       BY SIMILARITY.
FT   TRANSMEM       45     65       I (POTENTIAL).
FT   TRANSMEM      134    154       II (POTENTIAL).
FT   TRANSMEM      174    198       III (POTENTIAL).
FT   TRANSMEM      272    290       IV (POTENTIAL).
FT   TRANSMEM      333    353       V (POTENTIAL).
FT   TRANSMEM      372    398       VI (POTENTIAL).
FT   TRANSMEM      420    442       VII (POTENTIAL).
FT   TRANSMEM      523    541       VIII (POTENTIAL).
FT   TRANSMEM      581    602       IX (POTENTIAL).
FT   TRANSMEM      649    671       X (POTENTIAL).
FT   TRANSMEM      713    733       XI (POTENTIAL).
FT   METAL         565    565       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT   METAL         574    574       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ   SEQUENCE   740 AA;  82913 MW;  111422C2D33D30F0 CRC64;

Query Match             15.2%;  Score 60;  DB 1;  Length 740;
Best Local Similarity   48.1%;  Pred. No. 22;
  Matches   13;  Conservative    9;  Mismatches    1;  Indels    4;  Gaps    1;

Qy        47 WYTIFKDHVSLGDYEIHDGMNLELYQ 73
             ||  ||:.|::|.|||| | |  |  |
Db        21 WYAIAMAH---DFESHDGWTEENLYQ 43

RESULT 12
YHOB_DROME
ID   YHOB_DROME     STANDARD;      PRT;    644 AA.
AC   P12258;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   01-OCT-1989 (Rel. 12, Last annotation update)
DE   TRANSPOSABLE ELEMENT HOBO 108 73 KDA HYPOTHETICAL PROTEIN.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Streck R.D., Macgaffey J.E., Beckendorf S.K.;
RT   "The structure of hobo transposable elements and their insertion
RT   sites.";
RL   EMBO J. 5:3615-3623(1986).
```

```
DR   EMBL; X04705; CAA28410.1; -.
DR   PIR; A25684; A25684.
DR   FlyBase; FBgn0001210; hobo.
KW   Hypothetical protein.
SQ   SEQUENCE   644 AA;  73387 MW;  83168CE40F64D3A6 CRC64;

Query Match             15.1%;  Score 59.5;  DB 1;  Length 644;
Best Local Similarity   27.4%;  Pred. No. 22;
  Matches   17;  Conservative   10;  Mismatches   26;  Indels    9;  Gaps    2;

Qy        12 KKVRVKCNTDDTIGDLKKLIAAQTGTRWN-----KIVLKKWYTIFKDHVSLGDYEIHDG 65
             ||:  :   |: :   |   |:.:  |||        ||  |:.|||::||.||.|: ||
```

```
Db     340 KKIVKYCKKSNLQHTLETTLKSACPTRWNSNYKMMKSILDNWRSVDK---ILGEADIHVD 396

Qy      66 MN 67
              |
Db     397 FN 398


RESULT 13
ENV_MLVAV
ID   ENV_MLVAV       STANDARD;       PRT;    669 AA.
AC   P03386;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   ENV POLYPROTEIN PRECURSOR (COAT POLYPROTEIN) [CONTAINS: KNOB PROTEIN
DE   GP70; SPIKE PROTEIN P15E; R PROTEIN].
GN   ENV.
OS   AKV murine leukemia virus.
OC   Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses.
OX   NCBI_TaxID=11791;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84115072; PubMed=6319746;
RA   Herr W.;
RT   "Nucleotide sequence of AKV murine leukemia virus.";
RL   J. Virol. 49:471-478(1984).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=83090450; PubMed=6294621;
RA   Herr W., Corbin V., Gilbert W.;
RT   "Nucleotide sequence of the 3' half of AKV.";
RL   Nucleic Acids Res. 10:6931-6944(1982).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=82217034; PubMed=6283170;
RA   Lenz J., Crowther R., Straceski A., Haseltine W.;
RT   "Nucleotide sequence of the Akv env gene.";
RL   J. Virol. 42:519-529(1982).
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; J01998; AAB03092.1; -.
DR   EMBL; V01164; CAA24493.1; ALT_INIT.
DR   PIR; A03984; VCVWEK.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR002050; -.
DR   Pfam; PF00429; ENV_polyprotein; 1.
KW   Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.
FT   SIGNAL        1     31
FT   CHAIN        32    470       KNOB PROTEIN GP70.
FT   CHAIN       471    650       SPIKE PROTEIN P15E.
FT   CHAIN       651    669       R PROTEIN.
FT   TRANSMEM    473    493       POTENTIAL.
FT   TRANSMEM    612    632       POTENTIAL.
FT   CARBOHYD     43     43       N-LINKED (GLCNAC...). (POTENTIAL).
FT   CARBOHYD    199    199       N-LINKED (GLCNAC...). (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC...). (POTENTIAL).
FT   CARBOHYD    327    327       N-LINKED (GLCNAC...). (POTENTIAL).
FT   CARBOHYD    359    359       N-LINKED (GLCNAC...). (POTENTIAL).
FT   CARBOHYD    366    366       N-LINKED (GLCNAC...). (POTENTIAL).
FT   CARBOHYD    399    399       N-LINKED (GLCNAC...). (POTENTIAL).
FT   CONFLICT     35     35       G -> R (IN REF. 2).
FT   CONFLICT    463    463       E -> K (IN REF. 2).
FT   CONFLICT    592    592       E -> K (IN REF. 2).
SQ   SEQUENCE   669 AA;  73756 MW;  3A6C3845208A13F2 CRC64;
```

```
Qy       5 VCNDRLGKKVRVKCNTDDTIGDLKKLIAAQTGTRW 39
             :| : |:|:|||| |||| :||  |||| ||||
Db     385 LCIGAVPKTHQVLCNTTQKTSDGSYYLAAPTGTTW 419


RESULT 14
PSAB_SYNY3
ID   PSAB_SYNY3      STANDARD;       PRT;    730 AA.
AC   P29255; P73398;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PHOTOSYSTEM I P700 CHLOROPHYLL A APOPROTEIN A2.
GN   PSAB OR SLR1835.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92032774; PubMed=1932686;
RA   Smart L.B., McIntosh L.;
RT   "Expression of photosynthesis genes in the cyanobacterium
RT   Synechocystis sp. PCC 6803: psaA-psaB and psbA transcripts accumulate
RT   in dark-grown cells.";
RL   Plant Mol. Biol. 17:959-971(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA   Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA   Yamada M., Yasuda M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
CC   -!- FUNCTION: PSAA ALONG WITH PSAB BIND P700, THE PRIMARY ELECTRON
CC       DONOR OF PHOTOSYSTEM I (PSI) AS WELL AS THE ELECTRON ACCEPTORS A0,
CC       A1, AND FX.
CC   -!- COFACTOR: P700 IS A CHLOROPHYL DIMER, A0 IS CHLOROPHYLL, A1 IS A
CC       PHYLLOQUINONE AND FX IS A 4FE-4S IRON-SULFUR CENTER.
CC   -!- SUBUNIT: HETERODIMER OF A1 AND A2 SUBUNITS WHICH BIND TOGETHER THE
CC       IRON-SULFUR CENTER FX.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X58825; CAA41630.1; -.
DR   EMBL; D90906; BAA17438.1; -.
DR   PIR; S18243; S18243.
DR   PIR; S19090; S19090.
DR   InterPro; IPR001280; -.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Chlorophyll; Photosystem I; Electron transport; Photosynthesis;
KW   Transmembrane; Iron-sulfur; 4Fe-4S.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   TRANSMEM     45     68       I (POTENTIAL).
FT   TRANSMEM    134    156       II (POTENTIAL).
```

```
Query Match                    15.0%;  Score 59;  DB 1;  Length 669;
Best Local Similarity          37.1%;  Pred. No. 26;
Matches   13;  Conservative    4;  Mismatches  18;  Indels  0;  Gaps  0;
```

```
FT   TRANSMEM    174    198       III (POTENTIAL).
FT   TRANSMEM    272    290       IV (POTENTIAL).
FT   TRANSMEM    327    350       V (POTENTIAL).
FT   TRANSMEM    366    392       VI (POTENTIAL).
FT   TRANSMEM    414    436       VII (POTENTIAL).
FT   TRANSMEM    513    531       VIII (POTENTIAL).
FT   TRANSMEM    571    592       IX (POTENTIAL).
FT   TRANSMEM    639    661       X (POTENTIAL).
FT   TRANSMEM    703    723       XI (POTENTIAL).
FT   METAL       555    555       IRON-SULFUR (4FE-4S).
FT   METAL       564    564       IRON-SULFUR (4FE-4S).
FT   CONFLICT    536    537       DA -> ES (IN REF 1).
SQ   SEQUENCE    730 AA;  81160 MW;  EC804AA9F80E1231 CRC64;

Query Match
Best Local Similarity   15.0%;       Score 59;  DB 1;  Length 730;
                        48.1%;  Conservative  1;  Mismatches  1;  Indels  9;  Gaps  1;
Matches   13;

Qy    47 WYTIFKDHVSLGDYEIHDGMNLELYYQ 73
         ||  |    |:||||| |  ||
Db    21 WYGIATAH----DFETHDGMTEENLYQ 43
```

```
RESULT 15
PSAB_SYNP2
ID   PSAB_SYNP2      STANDARD;       PRT;     732 AA.
AC   P17155;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   PHOTOSYSTEM I P700 CHLOROPHYLL A APOPROTEIN A2.
GN   PSAB.
OS   Synechococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=32049;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cantrell A., Bryant D.A.;
RT   "Molecular cloning and nucleotide sequence of the psaA and psaB genes
RT   of the cyanobacterium Synechococcus sp. PCC 7002.";
RL   Plant Mol. Biol. 9:453-468(1987).
CC   -!- FUNCTION: PSAA ALONG WITH PSAB BIND P700, THE PRIMARY ELECTRON
CC       DONOR OF PHOTOSYSTEM I (PSI) AS WELL AS THE ELECTRON ACCEPTORS A0,
CC       A1, AND FX.
CC   -!- COFACTOR: P700 IS A CHLOROPHYL DIMER. A0 IS CHLOROPHYL, A1 IS A
CC       PHYLLOQUINONE AND FX IS A 4FE-4S IRON-SULFUR CENTER.
CC   -!- SUBUNIT: HETERODIMER OF A1 AND A2 SUBUNITS WHICH BIND TOGETHER THE
CC       IRON-SULFUR CENTER FX.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M18165; AAA88634.1; -.
DR   PIR; S06902; S06902.
DR   InterPro; IPR001280; -.
DR   Pfam; PF00223; psaA_psaB; 1.
DR   PRINTS; PR00257; PHOTSYSPSAAB.
DR   PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW   Chlorophyll; Photosystem I; Electron transport; Photosynthesis;
KW   Transmembrane; Iron-sulfur; 4Fe-4S.
FT   INIT_MET      0      0
FT   TRANSMEM     45     68       I (POTENTIAL).
FT   TRANSMEM    134    156       II (POTENTIAL).
FT   TRANSMEM    174    198       III (POTENTIAL).
FT   TRANSMEM    272    290       IV (POTENTIAL).
FT   TRANSMEM    329    352       V (POTENTIAL).
FT   TRANSMEM    368    394       VI (POTENTIAL).
FT   TRANSMEM    416    438       VII (POTENTIAL).
FT   TRANSMEM    515    533       VIII (POTENTIAL).
FT   TRANSMEM    573    594       IX (POTENTIAL).
FT   TRANSMEM    641    663       X (POTENTIAL).
FT   TRANSMEM    705    725       XI (POTENTIAL).
FT   METAL       557    557       IRON-SULFUR (4FE-4S).
FT   METAL       566    566       IRON-SULFUR (4FE-4S).
SQ   SEQUENCE    732 AA;  81557 MW;  94473FBFDC200EA1 CRC64;

Query Match
Best Local Similarity   15.0%;       Score 59;  DB 1;  Length 732;
                        48.1%;  Conservative  1;  Mismatches  1;  Indels  9;  Gaps  1;
Matches   13;

Qy    47 WYTIFKDHVSLGDYEIHDGMNLELYYQ 73
         ||  |    |:||||| |  ||
Db    21 WYGIATAH----DFETHDGMTEENLYQ 43

Search completed: June 13, 2001, 20:27:53
Job time: 205 sec
```

```
                    GenCore.version 4.5
            Copyright (c) 1993 - 2000  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        June 13, 2001, 19:30:07 ; Search time 178.32 seconds
                                          (without alignments)
                                          1119.634 Million cell updates/sec

Title:         US-09-331-930A-1
Perfect score: 342
Sequence:      1 gttccaggagattacagctc........atttttgttaagtgtact 342

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

  rched:       677276 seqs, 29189651 residues
Total number of hits satisfying chosen parameters:        1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     N_Geneseq_0401:*
          1:   /SIDS6/gcgdata/geneseq/geneseqn/NA1980.DAT:*
          2:   /SIDS6/gcgdata/geneseq/geneseqn/NA1981.DAT:*
          3:   /SIDS6/gcgdata/geneseq/geneseqn/NA1982.DAT:*
          4:   /SIDS6/gcgdata/geneseq/geneseqn/NA1983.DAT:*
          5:   /SIDS6/gcgdata/geneseq/geneseqn/NA1984.DAT:*
          6:   /SIDS6/gcgdata/geneseq/geneseqn/NA1985.DAT:*
          7:   /SIDS6/gcgdata/geneseq/geneseqn/NA1986.DAT:*
          8:   /SIDS6/gcgdata/geneseq/geneseqn/NA1987.DAT:*
          9:   /SIDS6/gcgdata/geneseq/geneseqn/NA1988.DAT:*
         10:   /SIDS6/gcgdata/geneseq/geneseqn/NA1989.DAT:*
         11:   /SIDS6/gcgdata/geneseq/geneseqn/NA1990.DAT:*
         12:   /SIDS6/gcgdata/geneseq/geneseqn/NA1991.DAT:*
         13:   /SIDS6/gcgdata/geneseq/geneseqn/NA1992.DAT:*
         14:   /SIDS6/gcgdata/geneseq/geneseqn/NA1993.DAT:*
         15:   /SIDS6/gcgdata/geneseq/geneseqn/NA1994.DAT:*
         16:   /SIDS6/gcgdata/geneseq/geneseqn/NA1995.DAT:*
         17:   /SIDS6/gcgdata/geneseq/geneseqn/NA1996.DAT:*
         18:   /SIDS6/gcgdata/geneseq/geneseqn/NA1997.DAT:*
         19:   /SIDS6/gcgdata/geneseq/geneseqn/NA1998.DAT:*
         20:   /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:*
         21:   /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:*
         22:   /SIDS6/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	342	100.0	342	20	X57359	P. obesus beacon D
2	342	100.0	342	21	C81767	Israeli sand rat b
3	235.2	68.8	419	20	X85745	Novel cDNA sequenc
4	205.6	60.1	413	21	C08654	Human secreted pro
5	205.6	60.1	427	21	C08655	Human secreted pro
6	197	57.6	502	21	C08656	Human secreted pro
7	184.2	53.9	333	16	T19132	Human gene signatu
8	180.4	52.7	315	16	T25014	Human gene signatu
9	123	36.0	624	21	C46615	Zea mays DNA fragm
10	118	34.5	556	21	C47007	Arabidopsis thalia
11	111.2	32.5	472	21	C41484	Zea mays DNA fragm

						Description
12	104.6	30.6	460	21	A31662	Plant microsatelli
13	77	22.5	102	20	X57370	Human beacon DNA.
14	77	22.5	102	21	C81770	Human beacon (shor
15	71	20.8	417	21	C08657	Human secreted pro
16	65.8	19.2	2180	21	C81769	Human beacon codin
17	64.2	18.8	311	21	A31226	Plant microsatelli
18	35.4	10.4	12492	8	N70543	Complete sequence
19	34.8	10.2	484	21	C41348	Zea mays DNA fragm
20	33.8	9.9	442	21	C36774	Arabidopsis thalia
21	33.8	9.9	501	21	C37049	Arabidopsis thalia
22	33.8	9.9	731	21	C43835	Arabidopsis thalia
23	33.8	9.9	976	21	C51382	Arabidopsis thalia
24	33.8	9.9	1106	21	C39920	Arabidopsis thalia
25	33.8	9.8	1462	21	C48056	Arabidopsis thalia
26	33.4	9.5	3083	21	C62751	Super ubiquitin pr
27	32.6	9.5	424	20	X84148	DNA encoding human
28	32.6	9.5	424	21	C79377	3' cDNA sequence o
29	32.4	9.5	903	21	C34429	Arabidopsis thalia
30	32.2	9.4	520	21	C41013	Zea mays DNA fragm
31	32.2	9.4	773	21	C41828	Arabidopsis thalia
32	32.2	9.4	3840	10	N92365	Sequence of maize
33	32	9.4	601	21	F11308	Aspergillus niger
34	31.8	9.3	1401	21	C44286	Arabidopsis thalia
35	31.8	9.3	1757	14	Q49428	Cytochrome p450 ho
36	31.8	9.3	1757	18	T94666	Petunia flavonoid
37	31.8	9.3	5174	20	Z28437	Sugar cane ubiquit
38	31.6	9.2	1087	21	C67872	Neospora caninum N
39	31.4	9.2	1965	21	C61108	Degenerate DNA seq
40	31	9.1	49999	20	Z23901	Human LOBO homolog
41	30.8	9.0	489	20	V86109	EST clone G1. Hom
42	30.6	8.9	1385	21	C51389	Arabidopsis thalia
43	30.4	8.9	1124	21	C44836	Arabidopsis thalia
44	30.4	8.9	2343	21	C61107	Human DNA encoding
45	30.4	8.9	5512	20	Z28435	Sugar cane ubiquit

```
                                                          ALIGNMENTS

RESULT   1
X57359
ID   X57359 standard; DNA; 342 BP.
XX
AC   X57359;
XX
DT   24-JUL-1999  (first entry)
XX
DE   P. obesus beacon DNA.
XX
KW   Beacon; hypothalamus; obese; lean; agonist; antagonist; treatment;
KW   obesity; anorexia; weight maintenance; energy imbalance; diabetes;
KW   metabolic syndrome; dyslipidemia; hypertension; insulin resistance;
KW   medicament; livestock; diagnosis; ss.
XX
OS   Psammomys obesus.
XX
FH   Key           Location/Qualifiers
FT   CDS           29..250
FT                 /*tag= a
FT                 /product= "beacon"
XX
PN   WO9923217-A1.
XX
PD   14-MAY-1999.
XX
PF   30-OCT-1998;  98WO-AU00902.
XX
PR   11-NOV-1997;  97AU-0000323.
PR   31-OCT-1997;  97AU-0000117.
XX
PA   (ITDI-) INT DIABETES INST.
PA   (UYDE-) UNIV DEAKIN.
XX
```

PI Collier G, Zimmet P2;
XX
DR WPI; 1999-337484/28.
DR P-PSDB; Y08413.
XX
PT New gene encoding a beacon protein associated with modulation of
PT obesity, diabetes and metabolic energy levels
XX
PS Claim 3; Page 49; 85pp; English.
XX
CC This invention describes a novel beacon protein and its encoding nucleic
CC acid which is expressed in larger amounts in hypothalamus tissue of obese
CC animals compared to lean animals. Agonists and antagonists of beacon can
CC be used to treat obesity, anorexia, weight maintenance, energy imbalance,
CC diabetes, metabolic syndrome, dyslipidemia, hypertension and/or insulin
CC resistance. The beacon protein, itself is used to manufacture medicaments
CC for treatment of obesity, anorexia, energy imbalance or diabetes. The
CC treatment is contemplated for both human and animals, such as those
CC important to the livestock industry. The antibody and polynucleotides are
CC useful in diagnosis of conditions as above.
XX
SQ Sequence 342 BP; 76 C; 79 G; 94 T; 0 other;

Query Match   100.0%; Score 342; DB 20; Length 342;
Best Local Similarity   100.0%; Pred. No. 2.9e-106;
Matches 342;   Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    1 gttccaggagattacagctccagccacaatgattgaggtggttgttgttgcaacgaccgtctagg  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 gttccaggagattacagctccagccacaatgattgaggtggttgttgttgcaacgaccgtctagg  60

Qy   61 aaagaaagtccgcgttaagtgcaactgcacaccgatgacaccatcgggagacttgaagaaactgat 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 aaagaaagtccgcgttaagtgcaactgcacaccgatgacaccatcgggagacttgaagaaactgat 120

Qy  121 agcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatttt 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 agcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatttt 180

Qy  181 taaggaccatgtatcctgggagattatgaaatccacgatgggatggaacctggagcttca 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 taaggaccatgtatcctgggagattatgaaatccacgatgggatggaacctggagcttca 240

Qy  241 ttaccagtagaggggaattcctccacctgcccaaccttgcttcctctcccatggctca 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ttaccagtagaggggaattcctccacctgcccaaccttgcttcctctcccatggctca 300

Qy  301 tttaacactgttgtagatgctcattttttgttaagtgtact 342
        |||||||||||||||||||||||||||||||||||||||||
Db  301 tttaacactgttgtagatgctcattttttgttaagtgtact 342
```

RESULT 2
C81767
ID C81767 standard; cDNA; 342 BP.
XX
AC C81767;
XX
DT 23-FEB-2001 (first entry)
XX
DE Israeli sand rat beacon coding sequence.
XX
KW Israeli sand rat; beacon; obesity; NIDDM; energy balance; diabetes;
KW ligand; ss.
XX
OS Psammomys obesus.
XX
PN WO20006493l-A1.
XX
PD 02-NOV-2000.
XX
PF 19-APR-2000; 2000WO-AU00342.

XX
PR 23-APR-1999;   99AU-0009919.
PR 24-MAR-2000;   2000AU-0006454.
XX
PA (AUTO-) AUTOGEN PTY LTD.
XX
PI Collier G, Walder K, Zimmet P;
XX
DR WPI; 2000-687311/67.
DR P-PSDB; B36290.
XX
PT Ligand of beacon protein useful for treating obesity, anorexia, energy
PT imbalance, diabetes, metabolic syndrome, dyslipidemia, hypertension and
PT insulin resistance -
XX
PS Claim 2; Fig 1; 67pp; English.
XX
CC The present invention is related to the isolation of a ligand known as
CC beacon from the Israeli sand rat. Beacon is associated with the
CC regulation of energy balance, and the protein, its coding sequence and
CC analogues can be used in the treatment of diabetes, obesity, anorexia,
CC energy imbalance, metabolic syndrome, dyslipidaemia, hypertension and
CC insulin resistance. In addition, they can be used in agriculture to
CC produce leaner animals.
XX
SQ Sequence 342 BP; 93 A; 76 C; 79 G; 94 T; 0 other;

Query Match   100.0%; Score 342; DB 21; Length 342;
Best Local Similarity   100.0%; Pred. No. 2.9e-106;
Matches 342;   Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    1 gttccaggagattacagctccagccacaatgattgaggtggttgttgttgcaacgaccgtctagg  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 gttccaggagattacagctccagccacaatgattgaggtggttgttgttgcaacgaccgtctagg  60

Qy   61 aaagaaagtccgcgttaagtgcaactgcacaccgatgacaccatcgggagacttgaagaaactgat 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 aaagaaagtccgcgttaagtgcaactgcacaccgatgacaccatcgggagacttgaagaaactgat 120

Qy  121 agcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatttt 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 agcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatttt 180

Qy  181 taaggaccatgtatcctgggagattatgaaatccacgatgggatggaacctggagctta 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 taaggaccatgtatcctgggagattatgaaatccacgatgggatggaacctggagctta 240

Qy  241 ttaccagtagaggggaattcctccacctgcccaaccttgcttcctctcccatggctca 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ttaccagtagaggggaattcctccacctgcccaaccttgcttcctctcccatggctca 300

Qy  301 tttaacactgttgtagatgctcattttttgttaagtgtact 342
        |||||||||||||||||||||||||||||||||||||||||
Db  301 tttaacactgttgtagatgctcattttttgttaagtgtact 342
```

RESULT 3
X85745
ID X85745 standard; cDNA; 419 BP.
XX
AC X85745;
XX
DT 06-SEP-1999 (first entry)
XX
DE Novel cDNA sequence from a mouse blastocyst cDNA library.
XX
KW Mouse; blastocyst; cDNA library; ss.
XX
OS Mus sp.
XX
PN JP1164691-A.
XX

PD 22-JUN-1999.
XX
PF 14-APR-1998; 98JP-0103115.
XX
PR 03-OCT-1997; 97JP-0271781.
XX
PA (RIKA) RIKAGAKU KENKYUSHO.
XX
PT New blastocyst cDNA - useful for library construction

Claim 3; Page 40; 41pp; Japanese.
XX
CC X85621-X85746 represent novel cDNA sequences that are isolated from a
CC mouse blastocyst cDNA library. The cDNA library was constructed from
CC C57B1/6 mice. The sequence can be used as a source of primers, probes
CC and complementary DNA sequences.
XX
SQ Sequence 419 BP; 104 A; 106 C; 102 G; 107 T; 0 other;

Query Match   68.8%;   Score 235.2; DB 20;   Length 419;
Best Local Similarity 87.8%;   Pred. No. 5.4e-70;
Matches 295; Conservative 0; Mismatches 28; Indels 13; Gaps 3;

Qy   1 gttccaggagattacagctccagccacaatgattgaggtggtttgcaacgaccgtctagg 60
         ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||| ||
Db   53 gtccaggagatattggagctccagcccacacaatgattgaggtggtttgcaacgaccgtctcgg 112

Qy   61 aagaaagtccgcgttaagtgcaacaccgatgacaccatcgggactgaagaaactgat 120
         ||||||||||||||||||| ||||||||||||||| ||||||||||||||||||
Db   113 aagaaaagtccgcgttaagt----caccgatgacaccatcggcgacttgaagaaactgat 168

Qy   121 acgggcccaaactggcactcgttggaataagatcgttccttaaaaagtggtacacgattt 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   169 agctgctccaaactggcacccggcgaaacaagatcgttccttaaaaagtggtacacgatttt 228

Qy   181 taaggaccatgtatctctgggagattatgaaatccacagatgggatgaacctggagcttta 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   229 taaggaccacgtgtctctggggagattatgaaatccacagatgggatgaacctggagcttta 288

Qy   241 ttaccagtagagggggaattcctccacct------tgcccaacttgctttcctctccc 292
         |||||||||||||||| |||||| ||| || ||| |||||||||||||||||
Db   289 ttaccagtagagggggggattcctctctctcctcgccctgcccctgccctcctctcccc 348

Qy   293 atggctcatttaacactgttgtagatgctcatttt 328
         ||||||||||||| ||||||||||||||||||||
Db   349 at-cctcatctgacactgtgtagatgtgtcatttt 383

RESULT 4
C08654
ID C08654 standard; cDNA; 413 BP.
XX
AC C08654;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 12729.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.

XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 12729; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 413 BP; 95 A; 103 C; 111 G; 98 T; 6 other;

Query Match   60.1%;   Score 205.6; DB 21;   Length 413;
Best Local Similarity 83.7%;   Pred. No. 5.9e-60;
Matches 246; Conservative 0; Mismatches 44; Indels 4; Gaps 1;

Qy   1 gttccaggagattacagctccagccacaatgattgaggtggttgcaacgaccgtctagg 60
         ||||||| || ||||||||||||||||||||||||||||||||||| |||||||||||||
Db   77 gttcgaggcgattcgagctccagctccagtgatcgagagttgttgcaacgaccgtctcggg 136

Qy   61 aagaaagtccgcgttaagtgcaacaccgatgacaccatcgggactgaagaaactgat 120
         |||||||| |||||||||||||||||| ||| |||||||||||| ||| | ||| ||||
Db   137 gagaaggtccgcgttaaatgcaacacggatgaccatcggggacctaagaagctgat 196

Qy   121 agcgggcccaaactggcactcgttggataagatcgttcctaaaaagtggtacacgatttt 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   197 tgcagcccaaactggtaccgttggaacagattgtcctgaacagattgtcctggagcttt 256

Qy   181 taaggaccatgtatctctgggagattatgaaatccacagatgggatgaacctggagcttta 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   257 taaggaccacgtgtctctgggggactatgaaatccacagtgggactgggatgaacctggagctta 316

Qy   241 ttaccagtagagggggaattcctccacttgctttcctctccccat 294
         ||| || ||| ||| ||| |||||||||||||||||||||||||
Db   317 ttatcaatagatgagaa----tccctcatcttcctgccccgctttcctcccccat 366

RESULT 5
C08655
ID C08655 standard; cDNA; 427 BP.
XX
AC C08655;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 12730.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX

PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 12730; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. NO ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 427 BP; 102 A; 103 C; 115 G; 106 T; 1 other;

Query Match   60.1%; Score 205.6; DB 21; Length 427;
Best Local Similarity   83.7%;   Pred. No.6e-60;
  Matches 246; Conservative   0; Mismatches   44; Indels   4; Gaps   1;

Qy   1 gttccaggagattacagctccagccacaatgattgaggtggttgcaacgaccgtccagg 60
          ||||-|| |||| ||||||||| | ||||||| |||| ||||||||||||||| ||
Db   77 gttcgaggcgattcgagctccagctcaggatgatcgaggttgttgcaacgaccgtctggg 136

Qy   61 aagaaaagtccgcgttaagtgcaacaccgatgacaccatcgggacttgaagaaactgat 120
          |||| ||||||||||| ||||||||| |||||||| ||||||||||-| ||||| |||||
Db   137 gaagaaggtccgcgttaaatgcaacagcaacacacgatatccatcgggacttaagaagtgat 196

Qy   121 agcggcccaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatttt 180
          |||||||||||||| || ||||||||||||| || |||||||||||||||||||||||||
Db   197 tgcagcccaaactggtacccgttggaacaagattgtcctgaagaagtggtacacgatttt 256

Qy   181 taaggaccatgtatctctggagagattatgaaatccacgatggatggaactggagcttta 240
          |||||||||| || |||||||||||| || |||||||||||||||||||||||||||||
Db   257 taaggaccacgtgtctctggggggactgtcctggggagcttta 316

Qy   241 ttaccagtagagggaattcctccacctgcccaacctgcttcctctcccat 294
          ||-|| ||||| || -||| ||| || |||||||||||||||||||||
Db   317 ttatcaatagatgagaa----tcctcatcctcctgcccgccctttcctctcccat 366

RESULT   6
C08656
ID   C08656 standard; cDNA; 502 BP.
XX
AC   C08656;
XX
DT   06-OCT-2000 (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 12731.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping: ss.

XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 12731; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 502 BP; 117 A; 122 C; 149 G; 114 T; 0 other;

Query Match   57.6%; Score 197; DB 21; Length 502;
Best Local Similarity   84.2%;   Pred. No. 5.3e-57;
  Matches 235; Conservative   0; Mismatches   40; Indels   4; Gaps   1;

Qy   16 agctccagccacaatgattgaggtggttgcaacgaccgcctaggaaagaaagtccgcgt 75
          ||||||||| |||| ||||||||||| |||| |||||||||||| || |||| |||||||
Db   172 agctccagctcaggatgatcgaggttgttgcaacgaccgtctggggagaaggtccgcgt 231

Qy   76 taagtgcaacaccgatgacaccatcgggacttgaagaaactgatagtagcggcccaaactgg 135
          || ||||||||| |||||| ||||||||||||| ||||| |||||| |||||||||||||||
Db   232 taaatgcaacagcgatgataccatcgggacttaagaagctgattgcagcccaaactgg 291

Qy   136 cactcgttgaataagtcgttcttaaaaagtggtacacgatcgtatc 195
          || ||||||| |||||| || ||||||||||||||||||||| || ||
Db   292 tacccgttgaacaagatgtcctgaagaagtggtacacggtcgtc 351

Qy   196 tctgggagattataaatccacgatgggatgaactggagctttattaccagtagagggg 255
          ||||||||| || |||||||||||||||| |||||||||||||| |||| ||-|| |||||
Db   352 tctggggactggaatccacgatggatggaacctggagcttatcaatagatgag 411

Qy   256 aattcctccacctgccaacctgcttcctctcccat 294
          || |||| |-||| || || |||||||||||||||
Db   412 aa----tcctcatcctcctgcccccgcttcctctcccat 446

RESULT   7
T19132
ID   T19132 standard; cDNA to mRNA; 333 BP.
XX
AC   T19132;
XX
DT   04-JUL-1996 (first entry)

XX
DE Human gene signature HUMGS00146.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
OS Homo sapiens.
XX
PN WO9514772-A1.
XX
PD 01-JUN-1995.
XX
PF 11-NOV-1994; 94WO-JP01916.
XX
PR 12-NOV-1993; 93JP-0355504.
XX
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
XX
PI Matsubara K, Okubo K;
XX
DR WPI; 1995-206931/27.
XX
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
PS Claim 1; Page 310; 2245pp; Japanese.
XX
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.

Sequence 333 BP; 91 A; 73 C; 76 G; 80 T; 13 other;

```
Query Match         53.9%;   Score 184.2;  DB 16;  Length 333;
Best Local Similarity  76.4%;   Pred. No. 9.7e-53;
Matches 233; Conservative  0;  Mismatches 68;  Indels  4;  Gaps

Qy   31 gattgaggtggttgcaacgaccgtctaggaagaaagtccgcgttaagtgcaacaccga 90
          || ||| |||||||||||||||| || || |||| |||| |||||| |||||| ||
Db    1 gatcgaggtgtttgcaacgaccgtctgggaagaaggtccgcgttaaatgcaacacgga 60

Qy   91 tgtaccacatcggggacttgaagaaactgatagcggccaaactggcactcgttggaataa 150
         || ||||||||||||| || ||||||||| || ||||||||||||| ||| ||||| ||
Db   61 tgtaccacatcggggacttgaagaaagctgattgcagccaaactggtacccgttggaacaa 120

Qy  151 gatcgttcctaaaaagtgtacacgatttttaaggacccatgtatctctggggagattatga 210
         ||| || ||||| || ||||||||||||| | |||||||||| | ||||||||||| ||
Db  121 gattgtcctgaagaagtgtacacgattttnaaggaccacgtgtctctgggggactatga 180

Qy  211 aatccacgatggggatgaacctgagctttattaccagtagaggggaattcctccacttg 270
         |||||||||||||| ||||||||||||| ||| ||| || | ||| ||| | || ||
Db  181 aatccacgatgggatgaacctggagcttattatcaatagatgagaa----tcctcatct 236

Qy  271 cccaaacctgctttcctctcccatggctcattaacactgttgtagatgctcattttttt 330
          |||| ||| ||||| ||| ||| ||||| |||| || ||| ||||| ||| ||
Db  237 tnctgcccgctntccnnctncccnctnccatcctccccacantnggatagatgcttngttt 296
```

```
Qy  331 gttaa 335
         || ||
Db  297 gtaaa 301
```

RESULT 8
T25014
ID T25014 standard; cDNA to mRNA; 315 BP.
XX
AC T25014;
XX
DT 07-NOV-1996 (first entry)
XX
DE Human gene signature HUMGS07136.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
OS Homo sapiens.
XX
PN WO9514772-A1.
XX
PD 01-JUN-1995.
XX
PF 11-NOV-1994;   94WO-JP01916.
XX
PR 12-NOV-1993;   93JP-0355504.
XX
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
XX
PI Matsubara K, Okubo K;
XX
DR WPI; 1995-206931/27.
XX
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
PS Claim 1; Page 1749; 2245pp; Japanese.
XX
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
SQ Sequence 315 BP; 73 A; 61 C; 83 G; 82 T; 16 other;

```
Query Match          52.7%;   Score 180.4;  DB 16;  Length 315;
Best Local Similarity  86.5%;   Pred. No. 1.8e-51;
Matches 199; Conservative  0;  Mismatches 31;  Indels  0;  Gaps  0;

Qy   31 gattgaggtggttgcaacgaccgtctaggaagaaagtccgcgttaagtgcaacaccga 90
          ||| ||||||||||||||||| || || |||| |||| |||||| |||||| ||
Db    1 gatcgaggtgtttgcaacgaccgtctgggaagaaggtccgcgttaaatgcaacacgga 60

Qy   91 tgtaccacatcggggacttgaagaaactgatagcggccaaactggcactcgttggaataa 150
```

```
Db    61  tgataccatcgggaccttaagaagctgattgcagcccaaactggtacccgttggaacaa  120
          |||||||||  || ||  |||||||||||||||||  ||||||||  |||||||  ||||
Qy   151  gatcgttcttaaaaagtggtacacgatttttaaggaccatgtatctctggagattatga  210
          || || ||  || || ||||||||||||||||||  ||||||  ||||||| || ||||
Db   121  gattgtcctgaagaagtggtacacgatttttggggaccacgtgtctctgggggactatga  180

Qy   211  aatccacgatgggatgaacctggagcttattaccagtagaggggaattc  260
          ||||||||||||||||||||||||||||||||| ||| ||| || ||||
Db   181  aatccacgatgggatgaacctggagcttattatcatagatgagaatgc   230
```

RESULT   9
C46615

ID   C46615 standard; DNA; 624 BP.
XX
AC   C46615;
XX
XX
DT   18-OCT-2000 (first entry)
XX
DE   Zea mays DNA fragment SEQ ID NO: 50789.
XX
XX
     Hybridisation assay; genetic mapping; gene expression control;
     protein identification; signal transduction pathway; metabolic;
     pathway; promoter; termination sequence; corn; ss.
XX
OS   Zea mays subsp. mays.
XX
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX

PR	Date	Accession		PR	Date	Accession
PR	25-FEB-1999;	99US-0121825.		PR	10-JUN-1999;	99US-0138540.
PR	05-MAR-1999;	99US-0123180.		PR	10-JUN-1999;	99US-0138847.
PR	09-MAR-1999;	99US-0123548.		PR	14-JUN-1999;	99US-0139119.
PR	23-MAR-1999;	99US-0125788.		PR	16-JUN-1999;	99US-0139452.
PR	25-MAR-1999;	99US-0126264.		PR	16-JUN-1999;	99US-0139453.
PR	29-MAR-1999;	99US-0126785.		PR	17-JUN-1999;	99US-0139492.
PR	01-APR-1999;	99US-0127462.		PR	18-JUN-1999;	99US-0139454.
PR	06-APR-1999;	99US-0128234.		PR	18-JUN-1999;	99US-0139455.
PR	08-APR-1999;	99US-0128714.		PR	18-JUN-1999;	99US-0139456.
PR	16-APR-1999;	99US-0129845.		PR	18-JUN-1999;	99US-0139457.
PR	19-APR-1999;	99US-0130077.		PR	18-JUN-1999;	99US-0139458.
PR	21-APR-1999;	99US-0130449.		PR	18-JUN-1999;	99US-0139459.
PR	23-APR-1999;	99US-0130510.		PR	18-JUN-1999;	99US-0139460.
PR	23-APR-1999;	99US-0130891.		PR	18-JUN-1999;	99US-0139461.
PR	28-APR-1999;	99US-0131449.		PR	18-JUN-1999;	99US-0139462.
PR	30-APR-1999;	99US-0133048.		PR	18-JUN-1999;	99US-0139463.
PR	30-APR-1999;	99US-0132407.		PR	18-JUN-1999;	99US-0139750.
PR	04-MAY-1999;	99US-0132484.		PR	18-JUN-1999;	99US-0139763.
PR	05-MAY-1999;	99US-0132485.		PR	21-JUN-1999;	99US-0139817.
PR	06-MAY-1999;	99US-0132486.		PR	22-JUN-1999;	99US-0139899.
PR	06-MAY-1999;	99US-0132487.		PR	23-JUN-1999;	99US-0140353.
PR	07-MAY-1999;	99US-0132863.		PR	23-JUN-1999;	99US-0140354.
PR	11-MAY-1999;	99US-0134256.		PR	24-JUN-1999;	99US-0140695.
PR	14-MAY-1999;	99US-0134218.		PR	28-JUN-1999;	99US-0140823.
PR	14-MAY-1999;	99US-0134219.		PR	29-JUN-1999;	99US-0140991.
PR	14-MAY-1999;	99US-0134221.		PR	30-JUN-1999;	99US-0141287.
PR	18-MAY-1999;	99US-0134370.		PR	01-JUL-1999;	99US-0141842.
PR	19-MAY-1999;	99US-0134768.		PR	01-JUL-1999;	99US-0142154.
PR	20-MAY-1999;	99US-0134941.		PR	02-JUL-1999;	99US-0142055.
PR	21-MAY-1999;	99US-0135124.		PR	06-JUL-1999;	99US-0142390.
PR	24-MAY-1999;	99US-0135353.		PR	08-JUL-1999;	99US-0142803.
PR	25-MAY-1999;	99US-0135629.		PR	09-JUL-1999;	99US-0142920.
PR	27-MAY-1999;	99US-0136021.		PR	12-JUL-1999;	99US-0142977.
PR	28-MAY-1999;	99US-0136392.		PR	13-JUL-1999;	99US-0143542.
PR	01-JUN-1999;	99US-0136782.		PR	14-JUL-1999;	99US-0143524.
PR	03-JUN-1999;	99US-0137222.		PR	15-JUL-1999;	99US-0144005.
PR	04-JUN-1999;	99US-0137528.		PR	16-JUL-1999;	99US-0144085.
PR	07-JUN-1999;	99US-0137502.		PR	16-JUL-1999;	99US-0144086.
PR	08-JUN-1999;	99US-0137724.		PR	19-JUL-1999;	99US-0144325.
PR	08-JUN-1999;	99US-0138094.		PR	19-JUL-1999;	99US-0144331.
				PR	19-JUL-1999;	99US-0144332.
				PR	19-JUL-1999;	99US-0144333.
				PR	19-JUL-1999;	99US-0144334.
				PR	19-JUL-1999;	99US-0144335.
				PR	20-JUL-1999;	99US-0144352.
				PR	20-JUL-1999;	99US-0144632.
				PR	20-JUL-1999;	99US-0144884.
				PR	21-JUL-1999;	99US-0144814.
				PR	21-JUL-1999;	99US-0144886.
				PR	21-JUL-1999;	99US-0145086.
				PR	22-JUL-1999;	99US-0145088.
				PR	22-JUL-1999;	99US-0145085.
				PR	22-JUL-1999;	99US-0145087.
				PR	22-JUL-1999;	99US-0145089.
				PR	22-JUL-1999;	99US-0145192.
				PR	23-JUL-1999;	99US-0145145.
				PR	23-JUL-1999;	99US-0145218.
				PR	23-JUL-1999;	99US-0145224.
				PR	26-JUL-1999;	99US-0145276.
				PR	27-JUL-1999;	99US-0145913.
				PR	27-JUL-1999;	99US-0145918.
				PR	28-JUL-1999;	99US-0145919.
				PR	28-JUL-1999;	99US-0145951.
				PR	02-AUG-1999;	99US-0146386.
				PR	02-AUG-1999;	99US-0146388.
				PR	03-AUG-1999;	99US-0146389.
				PR	04-AUG-1999;	99US-0147038.
				PR	04-AUG-1999;	99US-0147204.
				PR	05-AUG-1999;	99US-0147302.
				PR	05-AUG-1999;	99US-0147192.
				PR	06-AUG-1999;	99US-0147260.
				PR	06-AUG-1999;	99US-0147303.
				PR	06-AUG-1999;	99US-0147416.
				PR	09-AUG-1999;	99US-0147493.

```
         ||  |  ||||| ||||||||      |||||||| || || |||| |||| |||||
Db  127 gcgaagatgatcgagtggtgtgctgaacgaccgcctgggaagaagtgcgggtgaagtgc 186
         ||   |||||  ||||||||||||| ||  |||||| |||| ||| |||| || |||
Qy   83 aacaccgatgaccatcggggacctgaagaactgatagcggcccaaactgccactcgt 142
         ||  |||||||||||||| |||||||| || ||| |||||| || ||||| || |||
Db  187 aacgaggacgacaccatcgcgacctcagaaagctggtcggcggcgagacggggacgcgc 246
         |   |||||| | ||||||||||||| || ||||| ||||| ||||| |||| |||
Qy  143 tggaataagatcgttcttaaaagtggtacacgatttttaaggacaccatgtatcctggga 202
         |  |||||| ||||||||||||||| ||||| ||  ||||||||||||| || |||
Db  247 cccgaggagatccgcatccagaagtggtacaacatctacaaggacaccacatcacgctcaag 306
         |  |||||| ||||||||||||||| || |||||||| ||| || |||
Qy  203 gattatgaaatccacgatgggatgaacctggagcttattaccagtagagggggaattcct 262
         || || |||||||| |||||||| || ||| |||| || |||
Db  307 gactacgagatccacgacgggatggggctcgagctcgagctcctactacaactaggcgccctgaccc 366
Qy  263 ccaccttgccc 273
         |  ||| |||||
Db  367 caaccgtgccc 377


RESULT 10
C47007
ID   C47007 standard; DNA; 556 BP.
XX
AC   C47007;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 52220.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   30-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   06-MAY-1999; 99US-0132487.
PR   07-MAY-1999; 99US-0132863.
PR   11-MAY-1999; 99US-0134256.
PR   14-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   14-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   20-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135353.


PR   09-AUG-1999; 99US-0147935.
PR   10-AUG-1999; 99US-0148171.
PR   11-AUG-1999; 99US-0148319.
PR   12-AUG-1999; 99US-0148341.
PR   13-AUG-1999; 99US-0148565.
PR   13-AUG-1999; 99US-0148684.
PR   16-AUG-1999; 99US-0149368.
PR   17-AUG-1999; 99US-0149175.
PR   18-AUG-1999; 99US-0149426.
PR   20-AUG-1999; 99US-0149722.
PR   20-AUG-1999; 99US-0149723.
PR   20-AUG-1999; 99US-0149929.
PR   23-AUG-1999; 99US-0149902.
PR   23-AUG-1999; 99US-0149930.
PR   25-AUG-1999; 99US-0150566.
PR   26-AUG-1999; 99US-0150884.
PR   27-AUG-1999; 99US-0151065.
PR   27-AUG-1999; 99US-0151066.
PR   27-AUG-1999; 99US-0151080.
PR   30-AUG-1999; 99US-0151303.
PR   31-AUG-1999; 99US-0151438.
PR   01-SEP-1999; 99US-0151930.
PR   07-SEP-1999; 99US-0152363.
PR   10-SEP-1999; 99US-0153070.
PR   13-SEP-1999; 99US-0153758.
PR   15-SEP-1999; 99US-0154018.
PR   16-SEP-1999; 99US-0154039.
PR   20-SEP-1999; 99US-0154779.
PR   22-SEP-1999; 99US-0155139.
PR   23-SEP-1999; 99US-0155486.
PR   24-SEP-1999; 99US-0155659.
PR   28-SEP-1999; 99US-0156458.
PR   29-SEP-1999; 99US-0156596.
PR   04-OCT-1999; 99US-0157117.
PR   05-OCT-1999; 99US-0157753.
PR   06-OCT-1999; 99US-0157865.
PR   07-OCT-1999; 99US-0158029.
PR   08-OCT-1999; 99US-0158232.
PR   12-OCT-1999; 99US-0158369.
PR   13-OCT-1999; 99US-0159293.
PR   13-OCT-1999; 99US-0159294.
PR   13-OCT-1999; 99US-0159295.
PR   14-OCT-1999; 99US-0159329.
PR   14-OCT-1999; 99US-0159330.
PR   14-OCT-1999; 99US-0159331.
PR   14-OCT-1999; 99US-0159637.
PR   14-OCT-1999; 99US-0159638.
PR   18-OCT-1999; 99US-0159584.
PR   21-OCT-1999; 99US-0160741.
PR   21-OCT-1999; 99US-0160767.
PR   21-OCT-1999; 99US-0160768.
PR   21-OCT-1999; 99US-0160770.
PR   21-OCT-1999; 99US-0160814.
PR   22-OCT-1999; 99US-0160815.
PR   22-OCT-1999; 99US-0160980.
PR   22-OCT-1999; 99US-0160981.
PR   25-OCT-1999; 99US-0160989.
PR   25-OCT-1999; 99US-0161404.
PR   26-OCT-1999; 99US-0161405.
PR   26-OCT-1999; 99US-0161406.
PR   26-OCT-1999; 99US-0161359.
PR   26-OCT-1999; 99US-0161360.
PR   28-OCT-1999; 99US-0161361.
PR   28-OCT-1999; 99US-0161920.
PR   28-OCT-1999; 99US-0161992.
PR   29-OCT-1999; 99US-0161993.
PR   29-OCT-1999; 99US-0162142.


Query Match      36.0%;  Score 123;  DB 21;  Length 624;
Best Local Similarity  68.1%;  Pred. No. 7.5e-32;
Matches 171;  Conservative 0;  Mismatches 80;  Indels 0;  Gaps 0;

Qy   23 gccacaatgattgaggtggtttgcaacgaccgtctaggaagaaagtccgcgttaagtgc 82
```

us-09-331-930a-1.rng

PR	24-MAY-1999;	99US-0135529.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	21-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	22-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	23-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	24-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	28-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	29-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	30-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	01-JUL-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	02-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	06-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	14-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	21-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	25-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	26-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145551.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146336.	PR	26-OCT-1999;	99US-0161161.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.

```
PR   28-OCT-1999;  99US-0161992.
PR   28-OCT-1999;  99US-0161993.
PR   29-OCT-1999;  99US-0162142.

Query Match        34.5%;     Score 118;  DB 21;  Length 556;
Best Local Similarity  70.7%;    Pred. No. 3.5e-30;
Matches  157;  Conservative  0;  Mismatches  65;  Indels  0;  Gaps  0;

Qy   29 atgattgaggtggttcaacgaccgctctaggaagaaagtccgcgttaagtgcaacacc 88
         ||||| ||||||||| |||||| ||| |||| |||||| || |||| || ||||| |||
Db   70 atgatcgaggtggttcctcaacgatcgtttaggggaaaaagttaggtgaagtgtaacgat 129

Qy   89 gatgacaccatcggggacttgaagaaactgatagcgcccaaactgcactcgttgaat 148
         |||||||| || |||||| ||||||| ||||| |||||||| ||||| ||||| ||||
Db  130 gatgacacgatcgtgatctgaagaagctgtcgcggcacaaaccggaacacgagccgag 189

Qy  149 aagatcgttctaaaaagtggtacacgatttttaaggaccatgtatctctgggagattat 208
         ||||| ||| ||||||||||||| || ||||| || |||| ||| |||| | ||| |||
Db  190 aagatcagaattcagaagtggtacacactctacaaggatcacatcactctcaaggactat 249

Qy  209 gaaatcccacgatggatgaacctggagctttattaccagtag 250
         || |||| || ||| |||||||||| |||| ||| |||||
Db  250 gagatccatgacgcggcatggtcttgagcttcactacacaactag 291


RESULT  11
C41484
ID   C41484 standard; DNA; 472 BP.
XX
AC   C41484;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Zea mays DNA fragment SEQ ID NO: 32052.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway; metabolic;
KW   pathway; promoter; termination sequence; corn; ss.
XX
OS   Zea mays subsp. mays.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
PR   28-APR-1999;  99US-0130891.
PR   28-APR-1999;  99US-0131449.
PR   30-APR-1999;  99US-0132048.
PR   04-MAY-1999;  99US-0132407.
PR   05-MAY-1999;  99US-0132484.
PR   05-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   06-MAY-1999;  99US-0132487.
PR   07-MAY-1999;  99US-0132863.
PR   11-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
PR   14-MAY-1999;  99US-0134219.
PR   14-MAY-1999;  99US-0134221.

PR   14-MAY-1999;  99US-0134370.
PR   18-MAY-1999;  99US-0134768.
PR   19-MAY-1999;  99US-0134941.
PR   20-MAY-1999;  99US-0135124.
PR   21-MAY-1999;  99US-0135353.
PR   24-MAY-1999;  99US-0135629.
PR   25-MAY-1999;  99US-0136021.
PR   27-MAY-1999;  99US-0136392.
PR   28-MAY-1999;  99US-0136782.
PR   01-JUN-1999;  99US-0137222.
PR   03-JUN-1999;  99US-0137528.
PR   04-JUN-1999;  99US-0137502.
PR   07-JUN-1999;  99US-0137724.
PR   08-JUN-1999;  99US-0138094.
PR   10-JUN-1999;  99US-0138540.
PR   10-JUN-1999;  99US-0138847.
PR   14-JUN-1999;  99US-0139119.
PR   16-JUN-1999;  99US-0139452.
PR   16-JUN-1999;  99US-0139453.
PR   17-JUN-1999;  99US-0139492.
PR   18-JUN-1999;  99US-0139454.
PR   18-JUN-1999;  99US-0139455.
PR   18-JUN-1999;  99US-0139456.
PR   18-JUN-1999;  99US-0139457.
PR   18-JUN-1999;  99US-0139458.
PR   18-JUN-1999;  99US-0139459.
PR   18-JUN-1999;  99US-0139460.
PR   18-JUN-1999;  99US-0139461.
PR   18-JUN-1999;  99US-0139462.
PR   18-JUN-1999;  99US-0139463.
PR   18-JUN-1999;  99US-0139750.
PR   18-JUN-1999;  99US-0139763.
PR   21-JUN-1999;  99US-0139817.
PR   22-JUN-1999;  99US-0139899.
PR   23-JUN-1999;  99US-0140353.
PR   23-JUN-1999;  99US-0140354.
PR   24-JUN-1999;  99US-0140695.
PR   28-JUN-1999;  99US-0140823.
PR   29-JUN-1999;  99US-0140991.
PR   30-JUN-1999;  99US-0141287.
PR   01-JUL-1999;  99US-0141842.
PR   01-JUL-1999;  99US-0142154.
PR   02-JUL-1999;  99US-0142055.
PR   06-JUL-1999;  99US-0142390.
PR   08-JUL-1999;  99US-0142803.
PR   09-JUL-1999;  99US-0142920.
PR   12-JUL-1999;  99US-0142977.
PR   13-JUL-1999;  99US-0143542.
PR   14-JUL-1999;  99US-0143624.
PR   15-JUL-1999;  99US-0144005.
PR   16-JUL-1999;  99US-0144085.
PR   16-JUL-1999;  99US-0144086.
PR   19-JUL-1999;  99US-0144325.
PR   19-JUL-1999;  99US-0144331.
PR   19-JUL-1999;  99US-0144332.
PR   19-JUL-1999;  99US-0144333.
PR   19-JUL-1999;  99US-0144334.
PR   19-JUL-1999;  99US-0144335.
PR   20-JUL-1999;  99US-0144352.
PR   20-JUL-1999;  99US-0144632.
PR   20-JUL-1999;  99US-0144884.
PR   21-JUL-1999;  99US-0144814.
PR   21-JUL-1999;  99US-0145086.
PR   21-JUL-1999;  99US-0145088.
PR   22-JUL-1999;  99US-0145085.
PR   22-JUL-1999;  99US-0145087.
PR   22-JUL-1999;  99US-0145089.
PR   22-JUL-1999;  99US-0145192.
PR   23-JUL-1999;  99US-0145145.
PR   23-JUL-1999;  99US-0145218.
PR   23-JUL-1999;  99US-0145224.
PR   26-JUL-1999;  99US-0145276.
PR   27-JUL-1999;  99US-0145913.
```

```
PR    27-JUL-1999;    99US-0145918.
PR    27-JUL-1999;    99US-0145919.
PR    28-JUL-1999;    99US-0145951.
PR    02-AUG-1999;    99US-0146386.
PR    02-AUG-1999;    99US-0146388.
PR    02-AUG-1999;    99US-0146389.
PR    03-AUG-1999;    99US-0147038.
PR    04-AUG-1999;    99US-0147204.
PR    04-AUG-1999;    99US-0147302.
PR    05-AUG-1999;    99US-0147192.
PR    05-AUG-1999;    99US-0147260.
PR    06-AUG-1999;    99US-0147303.
PR    06-AUG-1999;    99US-0147416.
PR    09-AUG-1999;    99US-0147493.
PR    09-AUG-1999;    99US-0147935.
PR    10-AUG-1999;    99US-0148171.
PR    11-AUG-1999;    99US-0148319.
PR    12-AUG-1999;    99US-0148341.
PR    13-AUG-1999;    99US-0148565.
PR    13-AUG-1999;    99US-0148684.
PR    16-AUG-1999;    99US-0149368.
PR    17-AUG-1999;    99US-0149175.
PR    18-AUG-1999;    99US-0149426.
PR    20-AUG-1999;    99US-0149722.
PR    20-AUG-1999;    99US-0149723.
PR    20-AUG-1999;    99US-0149929.
PR    23-AUG-1999;    99US-0149902.
PR    23-AUG-1999;    99US-0149930.
PR    25-AUG-1999;    99US-0150566.
PR    26-AUG-1999;    99US-0150084.
PR    27-AUG-1999;    99US-0151065.
PR    27-AUG-1999;    99US-0151066.
PR    27-AUG-1999;    99US-0151080.
PR    30-AUG-1999;    99US-0151303.
PR    31-AUG-1999;    99US-0151438.
PR    01-SEP-1999;    99US-0151930.
PR    07-SEP-1999;    99US-0152363.
PR    10-SEP-1999;    99US-0153070.
PR    13-SEP-1999;    99US-0153758.
PR    15-SEP-1999;    99US-0154018.
PR    16-SEP-1999;    99US-0154039.
PR    20-SEP-1999;    99US-0154779.
PR    22-SEP-1999;    99US-0155139.
PR    23-SEP-1999;    99US-0155486.
PR    24-SEP-1999;    99US-0155659.
PR    28-SEP-1999;    99US-0156458.
PR    29-SEP-1999;    99US-0156596.
PR    04-OCT-1999;    99US-0157117.
PR    05-OCT-1999;    99US-0157753.
PR    06-OCT-1999;    99US-0157865.
PR    07-OCT-1999;    99US-0158029.
PR    08-OCT-1999;    99US-0158232.
PR    12-OCT-1999;    99US-0158369.
PR    13-OCT-1999;    99US-0159293.
PR    13-OCT-1999;    99US-0159294.
PR    13-OCT-1999;    99US-0159295.
PR    14-OCT-1999;    99US-0159329.
PR    14-OCT-1999;    99US-0159330.
PR    14-OCT-1999;    99US-0159331.
PR    14-OCT-1999;    99US-0159637.
PR    14-OCT-1999;    99US-0159638.
PR    18-OCT-1999;    99US-0159584.
PR    21-OCT-1999;    99US-0160741.
PR    21-OCT-1999;    99US-0160767.
PR    21-OCT-1999;    99US-0160768.
PR    21-OCT-1999;    99US-0160770.
PR    21-OCT-1999;    99US-0160814.
PR    21-OCT-1999;    99US-0160815.
PR    22-OCT-1999;    99US-0160980.
PR    22-OCT-1999;    99US-0160981.
PR    22-OCT-1999;    99US-0160989.
PR    25-OCT-1999;    99US-0161404.
PR    25-OCT-1999;    99US-0161405.
```

```
PR    25-OCT-1999;    99US-0161406.
PR    26-OCT-1999;    99US-0161359.
PR    26-OCT-1999;    99US-0161360.
PR    26-OCT-1999;    99US-0161361.
PR    28-OCT-1999;    99US-0161920.
PR    28-OCT-1999;    99US-0161992.
PR    28-OCT-1999;    99US-0161993.
PR    29-OCT-1999;    99US-0162142.


Query Match              32.5%;    Score 111.2;  DB 21;  Length 472;
Best Local Similarity    69.1%;    Pred No. 6.6e-28;
Matches 152;  Conservative    0;  Mismatches  68;  Indels    0;  Gaps    0;

Qy      29 atgatgaggtggtttgcaacgaccgtctaggaaagaaagtccgcgttaagtgcaacacc 88
           |||| ||||||| ||| ||||| |||||||| || || |||| | |||||||||
Db     148 atgatcgaggtggtgctcaacgaccgcctgggaagaaggtgcgggtcaagtgcaacgag 207

Qy      89 gatgacacaccatcgggacttgaagaaactgatagcggcccaaactggcactcgttggaat 148
           || |||||||||||| || | ||| ||| ||||| || |||| || || ||| |
Db     208 gacgacaccatcggcgacctcaataagctggtgtggccggccgagacgggggacgcgccccgag 267

Qy     149 aagatcgttcttaaaaatggtacacgattttaaggaccatgtatctctgggagattat 208
           |||||| | | |||||||| ||  ||||||| ||  || |||
Db     268 aagatccgcatccagaagtggtacaacatctacaggactacacatcaccctcaaggactac 327

Qy     209 gaaatccacgatgggatgaacctggagctttattaccagt 248
           || |||||| ||| ||| |||| | ||| |||
Db     328 gaggtccacgacgacgcatggcctcgagctctactacaact 367


RESULT 12
A31662
ID   A31662 standard; DNA; 460 BP.
XX
AC   A31662;
XX
DT   05-JUL-2000 (first entry)
XX
DE   Plant microsatellite marker #623.
XX
KW   Plant microsatellite sequence; core repeat sequence; detection; probe;
KW   DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW   variety identification; genetic variability evaluation; primer; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO9967421-A1.
XX
PD   29-DEC-1999.
XX
PF   25-JUN-1999;    99WO-NZ00092.
XX
PR   25-JUN-1998;    98US-0105307.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Havukkala IJ,  Bloksberg LN,  Glenn M;
XX
DR   WPI; 2000-116958/10.
XX
PT   New plant microsatellite markers and associated flanking species for
PT   the detection of polymorphic genetic markers -
XX
PS   Claim 1; Page 260; 392pp; English.
XX
CC   Sequences A31040-A32093 represent novel plant microsatellite sequences
CC   and associated flanking species. The sequences comprise a central core
CC   repeat sequence, especially selected from the sequences A32094-A32096
CC   with left and right flanking sequences. The polynucleotide sequences
CC   can be used in the detection of DNA polymorphisms, in genome mapping,
CC   in physical mapping, in positional cloning of genes, in variety
CC   identification and in evaluation of genetic variability within and
```

CC   between plant tissues, populations, cultivars, species and species
CC   groups. They may also be used to design hybridization probes for
CC   oligonucleotide fingerprinting and library screening and to design
CC   primers for microsatellite-primed PCR. Microsatellite markers are
CC   useful to locate specific economically useful genes in plant genomes.
XX
SQ   Sequence 460 BP; 122 A; 116 C; 146 G; 76 T; 0 other;

Query Match   30.6%; Score 104.6; DB 21; Length 460;
Best Local Similarity   69.1%; Pred. No. 1.1e-25;
Matches 143; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy   20 cagcccaaatgattgaggtggtttgcaacgaccgtctaggaaagaaagtccgcgttaag 79
          | |||| ||||||| ||||||| ||||||| || |||||| || ||| ||
Db   227 caagcggaagatgatagaggtggtgctgaacgaccggctggggaagaaggtgaggtgaag 286

Qy   80 tgcaacaccgatgacaccatcggggacttgaagaaactgatagcggcccaaactggcact 139
          |||||| ||||||||||||||| ||| | |||||| ||| ||| || ||||
Db   287 tgcaacgaggacgacgacaccatcggcgacctcaagaagctggtggccgcccagacgggcacc 346

Qy   140 cgttggaataagatcgttcttaaaaaatggtacacgatttttaaggaccatgtatctctg 199
          || ||||| | | |||||||| || |||||||| ||
Db   347 cgcgccgacaagatccgaatccagaagtggtacaacatctacaggaccaccacatcaccctc 406

Qy   200 ggagattatgaaatccacgatgggatg 226
          ||||| ||| |||||| ||| |||
Db   407 aaggactacgaggtccacgacggcatg 433

RESULT 13
X57370
ID   X57370 standard; DNA; 102 BP.
XX
AC   X57370;
XX
DT   24-JUL-1999 (first entry)
XX
DE   Human beacon DNA.
XX
KW   Beacon; hypothalamus; obese; lean; agonist; antagonist; treatment;
KW   obesity; anorexia; weight maintenance; energy imbalance; diabetes;
KW   metabolic syndrome; dyslipidemia; hypertension; insulin resistance;
KW   medicament; livestock; diagnosis; human; ss.
XX
OS   Homo sapiens.
XX
PN   WO0923217-A1.
XX
PD   14-MAY-1999.
XX
PF   30-OCT-1998; 98WO-AU00902.
XX
PR   11-NOV-1997; 97AU-0000323.
PR. 31-OCT-1997; 97AU-0000117.
XX
PA   (ITDI-) INT DIABETES INST.
PA   (UVDE-) UNIV DEAKIN.
XX
PI   Collier G, Zimmet PZ;
XX
DR   WPI; 1999-337484/28.
DR   P-PSDB; Y08414.
XX
PT   New gene encoding a beacon protein associated with modulation of
PT   obesity, diabetes and metabolic energy levels
XX
PS   Claim 3; Page 53; 85pp; English.
XX
CC   This invention describes a novel beacon protein and its encoding nucleic
CC   acid which is expressed in larger amounts in hypothalamus tissue of obese
CC   animals compared to lean animals. Agonists and antagonists of beacon can
CC   be used to treat obesity, anorexia, weight maintenance, energy imbalance,

CC   diabetes, metabolic syndrome, dyslipidemia, hypertension and/or insulin
CC   resistance. The beacon protein, itself is used to manufacture medicaments
CC   for treatment of obesity, anorexia, energy imbalance or diabetes. The
CC   treatment is contemplated for both human and animals, such as those
CC   important to the livestock industry. The antibody and polynucleotides are
CC   useful in diagnosis of conditions as above.
XX
SQ   Sequence 102 BP; 30 A; 21 C; 28 G; 22 T; 1 other;

Query Match   22.5%; Score 77; DB 20; Length 102;
Best Local Similarity   84.3%; Pred. No. 1.3e-16;
Matches 86; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy   29 atgattgaggtggtttgcaacgaccgtctaggaaagaaagtccgcgttaagtgcaacacc 88
          ||||| |||| ||||||||||||||||||||| || |||||| ||||||| |||||||||
Db   1 atgatcgaggtgttgcaacgaccgcctgggggaaaaaggtccncgttaaatgcaacacg 60

Qy   89 gatgacaccatcgggqacttgaagaaactgatagcggcccaa 130
          ||||| |||||||||||| | ||||| ||||| || || ||
Db   61 gatgataccatcggggacccttaagaagctgattgcagcctaa 102

RESULT 14
C81770
ID   C81770 standard; DNA; 102 BP.
XX
AC   C81770;
XX
DT   23-FEB-2001 (first entry)
XX
DE   Human beacon (short form) coding sequence.
XX
KW   Israeli sand rat; beacon; obesity; NIDDM; energy balance; diabetes;
KW   ligand; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO20064931-A1.
XX
PD   02-NOV-2000.
XX
PF   19-APR-2000; 2000WO-AU00342.
XX
PR   23-APR-1999;   99AU-0009919.
PR   24-MAR-2000; 2000AU-0006454.
XX
PA   (AUTO-) AUTOGEN PTY LTD.
XX
PI   Collier G, Walder K, Zimmet P;
XX
DR   WPI; 2000-687311/67.
DR   P-PSDB; B36291.
XX
PT   Ligand of beacon protein useful for treating obesity, anorexia, energy
PT   imbalance, diabetes, metabolic syndrome, dyslipidaemia, hypertension and
PT   insulin resistance -
XX
PS   Claim 3; Fig 1B; 67pp; English.
XX
CC   The present invention is related to the isolation of a ligand known as
CC   beacon from the Israeli sand rat. Beacon is associated with the
CC   regulation of energy balance, and the protein, its coding sequence and
CC   analogues can be used in the treatment of diabetes, obesity, anorexia,
CC   energy imbalance, metabolic syndrome, dyslipidaemia, hypertension and
CC   insulin resistance. In addition, they can be used in agriculture to
CC   produce leaner animals.
CC   Note: This sequence is stated in the claims as being a protein sequence,
CC   rather than a nucleic acid.
XX
SQ   Sequence 102 BP; 30 A; 21 C; 28 G; 22 T; 1 other;

Query Match   22.5%;   Score 77;   DB 21; Length 102;
Best Local Similarity   84.3%;   Pred. No. 1.3e-16;
Matches   86;   Conservative   0;   Mismatches   16;   Indels   0;   Gaps   0;

Qy   29 atgattgaggtggttgcaacgaccgtctaggaagaaagaaagtccgcgttaagtgcaacacc 88
          |||| |||| |||||||||||||||| || ||||||| ||||||||| ||||||||
Db   1 atgatccgaggttgttgcacgaccgtctgggaaaaggtccncgttaaatgcaacacg 60

Qy   89 gatgacaccatcggggacttgaagaaactgatagcggcccaa 130
          ||||| |||||||||| | ||||| ||||| || ||
Db   61 gatgataccatcggggaccttaagaagctgattgcagcctaa 102

RESULT 15
C08657
ID   C08657 standard; cDNA; 417 BP.
XX
AC   C08657;
XX
DT   06-OCT-2000 (first entry)
XX
KW   Human secreted protein 5' EST, SEQ ID NO: 12732.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 12732; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 417 BP; 85 A; 116 C; 110 G; 100 T; 6 other;

Query Match   20.8%;   Score 71;   DB 21; Length 417;
Best Local Similarity   84.2%;   Pred. No. 2.7e-14;
Matches   80;   Conservative   0;   Mismatches   15;   Indels   0;   Gaps   0;

Qy   1 gttccaggagattacagctccagccacacaatgattgaggtggtttgcaacgaccgtctagg 60
          |||| ||| |||||||||||||| | |||||||| | ||||||| |||||||||||| ||
Db   77 gttcgaggcgattcgagctccagctaggatgatcgcgaggttgttgcaacgaccgtctggg 136

Qy   61 aaagaaagtccgcgttaagtgcaacaccgatgaca 95
          ||||| |||||||||| |||||||| |||||| |
Db   137 gaagaaggtccgcgttaaatgcaacacggatgaaa 171

Search completed: June 13, 2001, 20:24:15
Job time: 3248 sec